

UNIVERSITE DE
YAOUNDE I
ECOLE NATIONALE
SUPERIEURE
POLYTECHNIQUE
DEPARTEMENT DE GENIE



UNIVERSITY OF
YAOUNDE I
NATIONAL ADVANCED
SCHOOL OF ENGINEERING
DEPARTEMENT OF
COMPUTER ENGINEERING

Intégration d'un module de calibration des paramètres du modèle DafMod au logiciel DafSim

Mémoire de fin d'étude/Master of Engineering

Présenté et soutenu par

Alex Christian Frank Billang

En vue de l'obtention du

Diplôme d'Ingénieur de Conception de Génie Informatique

Encadreur académique :

Dr. GEORGES EDOUARD KOUAMOU

Chargé de cours à l'Université de Yaoundé I

Encadreur professionnel :

Dr. VIVIEN ROSSI

Modélisateur statisticien HDR-CIRAD

Devant le jury constitué de :

PRESIDENT : MARTIN KOM, Pr

RAPPORTEUR : Georges KOUAMOU, Dr

EXAMINATEUR : Etienne TAKOU, Dr

INVITE : Vivien ROSSI, HDR.



Intégration d'un module de calibration des paramètres du Modèle
DafMod au logiciel DafSim

« Les mathématiques consistent à prouver des choses évidentes par des moyens complexes. »

George Polya.

DEDICACES

En mémoire à ma mère NGO BJLANG BRIGITTE.

REMERCIEMENTS

Ce travail est l'aboutissement de nombreux efforts et sacrifices et n'aurait jamais été accompli sans l'aide et le soutien d'un certain nombre de personnes Je remercie très respectueusement :

- **Professeur Martin KOM**, qui a accepté de présider le jury de ma soutenance,
- **Docteur Georges KOUAMOU**, pour ses conseils, sa disponibilité et son esprit paternel qui depuis plus de trois ans m'ont aidé à grandir sur plusieurs plans de ma vie.
- **Docteur Etienne TAKOU**, l'examineur, c'est un privilège de voir mon travail examiné par l'enseignant qui m'a donné cours d'analyse réelle et mathématique de base.
- **Docteur Vivien ROSSI**, qui m'a accueilli dans son équipe de recherche, pour ses bons conseils en modélisation statistique, pour ses nombreuses explications sur la dynamique des forêts qui furent très nécessaires pour l'aboutissement de mon travail et sa constante présence qui m'a poussé à me surpasser.

Je remercie mon collègue de bureau **Samuel NYOBE** Doctorant en Mathématique pour toutes ces notions en mathématique et les moments de détente que nous avons eu ensembles. A **Danielle ONDOBO** pour son soutien et son attention maternelle envers moi, A **Joseph Marie DOMGUIA** pour son aide dans la programmation.

Merci au corps enseignants et administratif de l'Ecole Nationale Supérieure Polytechnique(ENSP) de Yaoundé, plus spécialement les enseignants chargés de cours et de travaux dirigés du Département de Mathématique et Science Physique (MSP).

Je dis également merci à toute la promotion 2017, en particulier à ceux du Génie Informatique, les années passées ensemble on fait de nous une famille, elles nous ont unis dans les partages de connaissances et d'idées, les moments de joie et de rigolade, les moments de tristesse, les débats, je pense particulièrement à : **BAYOI MICHEL** camarade de Guerre, à l'imminent Délégué de classe et Frère, **YOBA Rostand**, à mon amie **SAMO Ronny**, à **MBA MBOGNE**, à **FOKAM Steve**, à **NDJAMA Joy**, à **Marcelin OLAMA** pour son aide dans la reformulation du mémoire et à **MVONDO Djob** pour son aide dans la préparation des examens, à tous Merci. Merci aussi aux promotions adjacentes à la nôtre, plus particulièrement ma marraine **Reine WAFO**.

Merci mes amis et amies et particulièrement à **ANGOS Avelin**, **NGANGAM Benoit**, **BAKAN Pierre** et **EYOUM** ; qui m'ont toujours témoigné de l'amour par leurs présences et leurs encouragements. A mes chers amis du niveau 1, 2 et 3 de polytechnique pour leur soutien dans ma foi et dans mon travail.

Enfin j'exprime ma vive gratitude à tous, mon grand frère **Serge NYOGOG BILLANG** pour tous les efforts qu'il s'est donnée pour que j'arrive jusque-là et à sa chère et tendre épouse **Généviève BILOUNGA**, à mes grandes sœurs **Nadège BILLANG** pour son soutien et ses nombreuses prières et **Christelle NYOGOG** qui a toujours su veillé à ce que je ne manque de rien, à **EYATA OWONA Françoise** ma bien aimée ,à ma petite sœur dans la foi **Josiane KOUAM** pour son aide durant ces deux dernières années et la joie qu'elle me procure, à **KAMEN Phalone** une grande sœur que j'aime, à **Prisca MAKSO** pour son soutien morale et spirituel , à **Jinnette MANAKABA** pour ses prières

et les merveilleux moments passés ensemble et à **Cédric WACHE** pour les épreuves et évènements que nous avons traversés ensemble.

AVANT PROPOS

Ce travail représente le résultat de cinq mois de stage académique, stage ingénieur effectué dans un bureau du département du Génie Informatique de l'École Nationale Supérieure Polytechnique pour le compte du **CIRAD** (Centre International en Recherche en Agronomie pour le Développement). Le **CIRAD** est un établissement public à caractère industriel et commercial (EPIC) français créé en 1984 et spécialisé dans la recherche agronomique appliquée aux régions chaudes. Dans le cadre du projet **DynAffor** (Dynamique des Forêts d'Afrique central), le **CIRAD** est l'un des partenaires de recherche de l'enseignement. Ce projet vise à quantifier l'effet sur la dynamique forestière et sur les processus qui la pilotent, C'est un projet de plus de six millions d'euros, financé par le Fonds Français pour l'Environnement Mondial (FFEM) et l'Agence Française de Développement (AFD) à hauteur d'environ 2.6 millions d'euros. Il se déroule sur une première phase de 5 ans depuis 2012, sous l'égide de la COMIFAC qui regroupe cinq Pays d'Afrique Centrale : Cameroun, Congo, République Centrafricaine, République Démocratique du Congo. C'est par la collaboration entre le **CIRAD** et **UY1** (l'Université de Yaoundé I) que nous sommes parvenus à travailler sur ce projet.

Notre travail s'est focaliser sur le moteur du logiciel DafSim construit dans le cadre du projet **DynAffor**. Ce logiciel doit produire des simulations du comportement de la dynamique des forêts afin d'aider les exploitants forestiers sur les règles d'aménagement à prendre pour leur gestion durable en Afrique Centrale.

RESUME

Des chercheurs du CIRAD ont pu mettre au point un simulateur de la dynamique forestière utilisant un modèle calibré sur les données collectées depuis 1980 sur le dispositif expérimental de M’Baïki en République Centrafricaine. Mais ce simulateur à l’état où nous l’avons trouvé ne satisfait pas les objectifs du CIRAD car il ne prend pas en compte le cas d’inventaires de données collectées à des intervalles de temps irréguliers. Notre travail consiste à mettre en place un module qui permettra au logiciel d’intégrer la gestion des inventaires de données irrégulières afin d’améliorer la qualité des résultats de simulation.

Pour atteindre cet objectif, nous avons d’abord étudié le fonctionnement du simulateur afin de comprendre les entrées et les sorties, puis nous avons proposé une modélisation des processus directs de la dynamique forestière et enfin nous avons proposé un algorithme qui permet d’inférer sur l’évolution des arbres à partir de données collectées dans leurs environnements de manière régulière (chaque année) ou de manière irrégulière (plus d’une année après) suivi d’une analyse des résultats obtenus. L’implémentation s’est appuyée sur le logiciel de calcul statistique R utilisé aussi pour mettre en place le simulateur à travers ses nombreuses extensions dédiées au développement logiciel.

Mots clés : dynamique forestière ; processus de dynamique forestière ; modèle de dynamique forestière ; inventaires réguliers, inventaires irréguliers.

ABSTRACT

CIRAD researchers have been able to develop a forest dynamics simulator using a forest dynamics model calibrated on the data collected since 1980 on the experimental forest dynamics monitoring system of M'Baïki in the Central African Republic. But this simulator in the state in which we found it did not satisfy the objectives of CIRAD because it did not handle the case of inventories of data collected irregularly over a period that could vary. Our work thus consisted during this internship to put in place a module which will allow the software to take into account the case of the irregular data inventories in order to improve the quality of the results of simulation.

To achieve this objective, we first studied the functioning of the simulator in order to understand the inputs and the outputs, then we proposed a modeling of the processes director of the forest dynamics and finally we developed an algorithm which makes it possible to infer on The evolution of trees from data collected in their environments on a regular basis (each year) or irregularly (more than one year later) followed by an analysis of the results obtained. The implementation was based on the software of statistical calculation R also used to develop the simulator through its numerous extensions intended for the software development.

Key words: forest dynamics; Forest dynamics process; Forest dynamics model; Regular inventories, irregular inventories.

PRINCIPALES NOTATIONS

- X : l'ensemble des variables explicatives (partie déterministe)
- Y : la variable aléatoire à expliquer.
- L : comme indice, notation du nombre de variables explicatives.
- s : indice d'espèce d'arbre.
- S : nombre d'espèce.
- X_s : toutes les données de variables explicatives correspondant à l'espèce s .
- $x_{s,i} = (1, x_{s,i,1}, \dots, x_{s,i,L})$: données de variables explicatives de l'espèce s à la ligne i du tableau de données (selon son ordre).
- $x_{s,j} = (1, x_{s,j,1}, \dots, x_{s,j,L})$: j^e données de variables explicatives de l'espèce s .
- a : un arbre
- θ_{-n} : Symbole qui désigne l'absence des données de 'n'année (s)
- θ_{+n} : Symbole qui désigne la complétion de 'n'année(s) par interpolation
- $P(\lambda)$: loi de Poisson de paramètre λ .
- $B(n,p)$: loi Binomiale de paramètre n (nombre d'épreuve) et p (probabilité du succès).
- \mathbb{P} : Symbole de calcul de la probabilité d'un évènement.
- E : *symbole de calcul de l'espérance mathématique.*
- \ln : fonction logarithme népérienne
- \exp : fonction exponentielle.
- P : notation de la vraisemblance.
- L : Comme fonction, est la notation de la log-vraisemblance : logarithme de la vraisemblance.
- $\Theta_k = (\Theta_{k,0} \dots \Theta_{k,L})$: paramètre à estimer ;
- $\binom{n}{p}$: combinaison de p dans n : nombre de sous-ensembles d'un ensemble à n éléments qui contiennent p éléments.
- $\frac{\delta f}{\delta x}$: dérivée partielle de la fonction f par rapport à la variable x .

TABLE DES MATIERES

DEDICACES.....	5
REMERCIEMENTS	6
AVANT PROPOS.....	8
RESUME.....	9
ABSTRACT	10
PRINCIPALES NOTATIONS	11
TABLE DES MATIERES.....	12
LISTE DES FIGURES	14
LISTE DES TABLEAUX	15
GLOSSAIRE.....	16
INTRODUCTION GENERALE.....	17
Contexte	17
Problématique.....	18
Objectifs	18
Plan du mémoire.....	18
Chapitre 1 : Etat de l'art	20
Introduction.....	20
1.1. Etat Des lieux sur le simulateur d'exploitation forestière	20
1.1.1. Fonctionnement du Simulateur.....	20
1.1.2. Description du Modèle du Simulateur.....	21
1.2. Les Méthodes et outils utilisé	24
1.2.1. Les outils	24
1.2.2. Les méthodes	27
Récapitulatif	30
Chapitre 2 : Mise en œuvre	31
Introduction.....	31
2.1. Définition globale du problème	31
2.1.1. Modélisation pour les intervalles de Temps irréguliers entre les inventaires	32
2.1.2. Interpolation des données de variables explicatives	42
2.2. Méthodologie pour la Résolution du problème	43
2.2.1. Création du nouveau format de données	43
2.2.3. Méthode de validation des Résultats	45
Récapitulatif :	49

Chapiste 3 : Présentation et Analyse des Résultats	50
Introduction.....	50
3.1. Présentation des outils utilisés pour l'implémentation.....	50
3.1.2. Le logiciel R.....	50
3.2.2. Le logiciel RStudio.....	51
Récapitulatif :.....	51
3.3. Présentation des Résultats obtenus	52
3.3.1. Echantillon de test	52
3.3.2 Valeurs des paramètres estimés.....	54
Récapitulatif.....	59
3.3.3 Récapitulatif	61
Conclusion.....	62
BIBLIOGRAPHIE	63

LISTE DES FIGURES

Figure 1 Forêts Tropicales	17
Figure 2 Principe de fonctionnement du simulateur 'DAfSim' [Dyn14a]	20
Figure 3 Processus de la dynamique forestière du modèle 'DafSim'[ROS15]	22
Figure 4 Représentation de l'évolution du peuplement [ROS15]	22
Figure 5 Processus d'évolution d'un peuplement [Dyn14c].	23
Figure 6 Localisation des parcelles expérimentales de M'Baiki	24
Figure 7 description de la méthode Mote Carlo [Bon09]	27
Figure 8 DHP: diamètre à hauteur de poitrine (siteA)	28
Figure 9 Chaîne des opérations de l'inventaire forestier ([CMUC17])	31
Figure 10 Automate de croissance ou de mortalité pour d années (TEM2017)	37
Figure 11 organigrammes de l'algorithme	45
Figure 12 logo de R et de RStudio	50
Figure 13 Logo de RStudio	51
Figure 14 Erreur absolue pour un an	58
Figure 15 Erreur absolue pour deux ans	58
Figure 16 Erreur absolue pour trois ans	59
Figure 17 Comparaison des RMSE	60

LISTE DES TABLEAUX

Tableau 1 classe de diamètre	33
Tableau 2 Tableau de données pour les croissances, $i=1 \dots N$ ([TEM2017])	35
Tableau 3 pour le recrutement $i=1 \dots N$ ([TEM2017])	39
Tableau 4 première partie du tableau de donnée pour la classe [9-20]	53
Tableau 5 deuxième partie du tableau pour la classe [9-20]	53
Tableau 6 Données du paramètre θ initial	54
Tableau 7 paramètre θ pour moins un an.....	54
Tableau 8 paramètre θ pour l'année complété.....	54
Tableau 9 paramètre θ pour moins deux ans	55
Tableau 10 paramètre θ pour deux années complétés	55
Tableau 11 paramètre θ pour deux ans complétés	55
Tableau 12 paramètre θ pour trois années complétés	55
Tableau 13 Erreur absolue	57
Tableau 14 formule de la RMSE	59
Tableau 15 Tableau des RMSE	60
Tableau 16 Récapitulatif.....	61

GLOSSAIRE

Sigle	Définition
CIRAD	Centre Internationale de Recherche en Agronomie pour le Développement
DME	Diamètre Minimum d'Exploitation
DMA	Diamètre Minimum d'Aménagement

INTRODUCTION GENERALE

Contexte

L'Afrique possède plus au moins $\frac{3}{4}$ des régions intertropicales du monde. Par sa végétation, il est un important océan vert c'est-à-dire que : le continent est composé des forêts, de savanes boisées, de savanes herbeuses et des steppes. Son existence depuis toujours contribue, à sa mesure, à l'équilibre écologique de la planète. En effet, au cours de ces dernières décennies, la forêt tropicale est de plus en plus menacée. On constate au centre de l'Afrique, l'exploitation illégale et l'abattage abusif d'arbres : certaines essences d'arbres rares sont exportées à travers le monde. Le déboisement dû à la surexploitation de la forêt. L'utilisation intensive du charbon de bois appauvrissent la forêt. Tout ceci entraîne des conséquences écologiques, sociales et économiques : le réchauffement climatique et la perturbation de saison, la guerre civile brutale, les nombreux conflits tribaux. Plusieurs espèces sont en voie de disparition. Pour préserver la survie de cette abondance naturelle pleine d'avantages, nous avons donc besoin de comprendre, d'expliquer et de prédire le comportement de la forêt en particulier la forêt tropicale d'Afrique centrale. Partant du fait que le système d'évolution de la forêt est dynamique car la population forestière en milieu tropicale est hétérogène ; nous allons utiliser des modèles statistiques de dynamique des populations qui présentent des conventions adaptables et consistantes pour avoir une idée d'un tel peuplement.

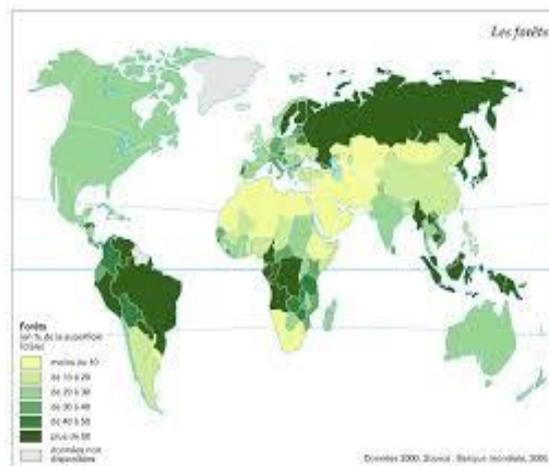


Figure 1 Forêts Tropicales

Problématique

La forêt d'Afrique Centrale a une très grande biodiversité. On n'y dénombre pas moins de 11 000 espèces de végétaux. C'est une biodiversité irremplaçable et également un atout majeur pour le pays et la planète en matière de lutte contre les changements climatiques. Nul n'ignore le rôle capital que joue la forêt tropicale dans la régularité climatique. Non seulement elle protège contre les intempéries mais aussi elle contribue à la survie de la population. Inquiet des conséquences de l'exploitation du bois d'œuvre de cette forêt les pays d'Afrique Centrale ont redéfinis les conventions forestières dans le but de trouver des moyens de produire durablement du bois d'œuvre. Ce qui implique la quantification des processus qui dirigent la dynamique forestière : le recrutement, la mortalité et la croissance. Ainsi le projet DynafFor : Dynamique des forêts d'Afrique Centrale, où se développe le logiciel DafSim dont le moteur développé à l'usage de flexmix (Bibliothèque du langage R) Mais ce dernier ne présente aucune considération particulière lorsque les données sont collectées à des intervalles de temps irréguliers (lorsque la période en année entre deux collectes de données est supérieure à 1). Comment donc mettre en place un module de calibration des paramètres du modèle DafMod de manière à ce que le moteur de simulation de la dynamique forestière prenne en compte l'irrégularité des inventaires de données sur le terrain ?

La réponse à cette interrogation passe par la réponse aux questions suivantes :

- Comment fonctionne le moteur de ce simulateur et qu'elles sont les données produites en sortie de ce simulateur ?
- Quels sont les processus directeurs de la dynamique forestière qui fournissent des informations pertinentes pour la gestion durable des forêts et comment modéliser ses processus?
- Comment s'assurer que les modèles choisis fournissent de très bons résultats et ainsi pouvoir affirmer que la simulation effectuée par DafSim sera meilleure avec ce module ?

Objectifs

Le thème soumis à notre étude nécessite qu'on modélise les processus directeurs de la dynamique forestière et par la suite écrire un algorithme pour mettre en œuvre ses modèles qui constitueront notre module et enfin développer le module de calibration des paramètres du modèle **DafMod** intégrable à DafSim.

Plan du mémoire

Ce travail est composé en trois grandes parties organisées comme suit :

1. La présentation de l'existant qui a pour principal objectif de présenter l'état des lieux, les concepts liés à notre étude et de recenser les éléments liés à la gestion forestière.

2. La mise en œuvre, où nous présentons notre modélisation de la solution dans le cas d'inventaires de données relevés de manière quelconque. Cette partie s'achève avec la production d'équations qui guideront le développement de l'algorithme qui constituera le cœur de notre module.
3. Enfin, nous terminerons avec la partie Test et présentation des résultats obtenus sur un échantillon collecté de manière régulière sur 10 ans dans lequel nous avons introduit graduellement des imperfections pour des sauts de 1an, 2ans etc...

Chapitre 1 : Etat de l'art

Introduction

Le simulateur appelé **DafSim** a été développé dans le cadre du projet **DynAffor**. Il a pour principal objectif de simuler l'évolution d'un peuplement forestier et est basé sur le modèle statistique **DafMod** qui a été calibré, en utilisant les données collectées sur le dispositif de **M'Baiki**. Ce simulateur a été réalisé à l'aide du langage de calcul statistique **R** [Fou14] et il est basé sur les méthodes de **Monte Carlo**. Nous allons donc présenter dans ce chapitre les différents concepts liés à notre étude.

1.1. Etat Des lieux sur le simulateur d'exploitation forestière

DafSim est le simulateur d'exploitation forestière développé dans le cadre du projet **DynAffor** (*DynAfforSimulateur*) pour explorer les réponses des forêts face aux différentes pratiques d'exploitation [Dyn14a].

1.1.1. Fonctionnement du Simulateur

1.1.1.1. Principe

Le simulateur **DafSim** est basé sur le modèle **DafMod**. Il présente donc certaines caractéristiques liées à ce choix : le peuplement, les groupes d'espèces ou l'espèce d'intérêt sont représentés par une distribution des effectifs d'arbres dans des classes de diamètre, sur une superficie donnée (celle de l'inventaire) [Dyn14a].

Le simulateur reproduit l'évolution du peuplement forestier avec un pas de temps fixe choisi (par exemple annuel : un an) en appliquant des exploitations à une périodicité fixée (par exemple tous les 30 ans) [Dyn14a].

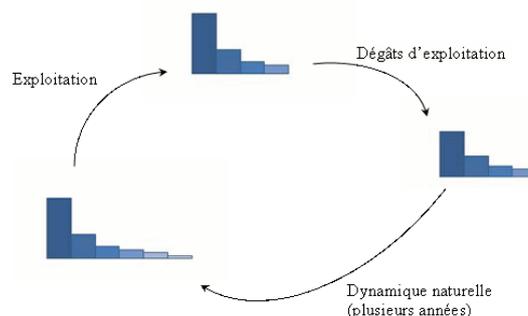


Figure 2 Principe de fonctionnement du simulateur 'DAfSim' [Dyn14a]

Le simulateur est composé de trois modules : exploitation, dégâts d'exploitation et dynamique naturelle. Chacun des modules agit sur le peuplement au moment déterminé par le simulateur : le module « exploitation » est exécuté périodiquement (par exemple tous les 30 ans), le module « dégâts d'exploitation » est exécuté juste après l'exploitation et le module « dynamique naturelle » est exécuté à tous les pas de temps [Dyn14a].

1.1.1.2. Données nécessaires

Plusieurs informations sont nécessaires pour être en mesure d'utiliser le simulateur :

1. Il faut avoir calibré le modèle **DafMod** avec des données issues de parcelles ou de sentiers, représentatifs du peuplement ou de la population à simuler.
2. Il faut avoir une représentation du peuplement ou de la population qui servira comme point de départ aux simulations. C'est-à-dire la distribution des effectifs par classe de diamètre du peuplement, du groupe d'espèces ou de l'espèce sur lesquels les simulations doivent être réalisées.
3. Il faut caractériser l'exploitation : liste des essences exploitées et de leur DMA¹ respectif, intensité de coupe, durée des rotations ...
4. Il faut définir la proportion d'arbres morts par classe de diamètre due aux dégâts d'exploitation [Dyn14a].

1.1.1.3. Sorties du simulateur

Le simulateur nous permet d'obtenir l'évolution, sur la durée de la simulation, de différents indicateurs avec leurs intervalles de confiance. Tout indicateur calculable à partir de la représentation du peuplement en classes diamétriques peut être obtenu, par exemple : la surface terrière pour le peuplement, le groupe d'espèces ou une espèce particulière, ou encore les proportions d'espèces selon leur tempérament [Dyn14a].

1.1.1.4. Validité des simulations

La confiance dans les simulations est d'autant plus grande que le nombre de données utilisées pour calibrer le modèle est élevé. En particulier si les données sont des suivies de long-terme sur des parcelles ayant subies des exploitations d'intensités contrastées. Elle est d'autant plus grande que les simulations n'amènent pas le peuplement dans un état éloigné de ceux des données de calibration [Dyn14a].

1.1.2. Description du Modèle du Simulateur

Le modèle de dynamique forestière utilisé pour la mise en place de ce simulateur est appelé **DynAfforMoDel** en abrégé **DafMod**. Il a été conçu avec pour principal objectif d'appréhender la réponse de la forêt face à l'exploitation. Nous vous présenterons ici les grands axes de ce modèle pour mieux l'expliquer [Dyn14b].

¹ Diamètre minimum d'Aménagement

1.1.2.1. La structure du modèle

La dynamique forestière est la résultante de trois processus : la régénération, la croissance et la mortalité. Le modèle de dynamique forestière développé dans le cadre du projet **DynAffFor** est également composé de trois sous-modèles, un pour chaque processus. Les sous-modèles sont autonomes, ce qui permet de les calibrer, de le modifier ou même d'en créer de nouveaux indépendamment des autres éléments du modèle. [Dyn14c].

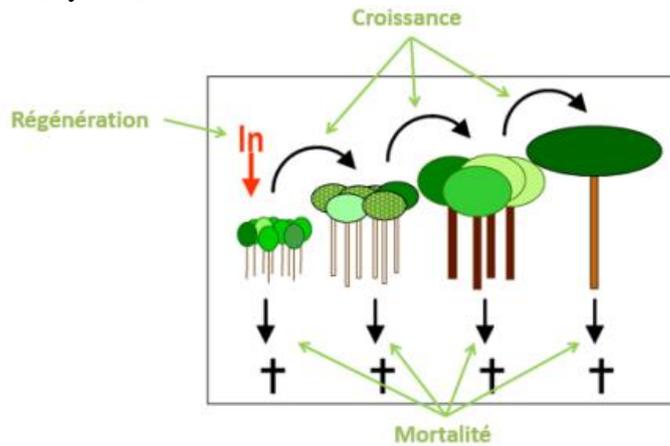


Figure 3 Processus de la dynamique forestière du modèle 'DafSim'[ROS15]

1.1.2.2. Représentation du peuplement forestier

Le peuplement est représenté par une distribution des effectifs d'arbres dans des classes de diamètre, sur une surface donnée. Cette distribution peut être calculée pour l'ensemble du peuplement, ou par groupe d'espèces, ou par espèce. Le modèle fait évoluer les arbres de chaque classe de diamètre (ils peuvent rester dans leur classe, passer dans la classe suivante ou bien mourir, sur un pas de temps donné). Techniquement, ce modèle fait partie de la famille des modèles matriciels fonctionnant à l'échelle de la communauté. [Dyn14c].

	Ayous	Azobé	Bilinga	...	Otunga	Sapelli
[10;20[40	15	5	...	50	26
[20;30[30	16	0	...	37	22
...
[110;120[3	0	1	...	1	8
[120; +∞[0	0	0	...	0	2

Figure 4 Représentation de l'évolution du peuplement [ROS15]

1.1.2.3. Prise en compte de l'exploitation forestière

Pour que le modèle soit en mesure de reproduire aussi bien la dynamique post-exploitation que la dynamique des forêts non perturbées, des variables caractérisant la structure du peuplement (par exemple la surface terrière des classes diamétriques ou le nombre d'arbre par classe de diamètre) sont intégrées dans les sous-modèles des processus. [Dyn14c].

1.1.2.4. Calibration du modèle

Des inventaires réguliers (au moins deux) sont nécessaires pour produire des données permettant de calibrer le modèle. Pour que le modèle soit en mesure de reproduire le comportement de la forêt non perturbée, il faut suivre des arbres en situation non perturbée. Pour qu'il soit en mesure de reproduire le comportement de la forêt après exploitation, il faut suivre des arbres après exploitation. Ceci explique la configuration des sentiers et des parcelles intégrés dans **DynAffor**, comportant autant que possible des sites intacts et des sites exploités. [Dyn14c].

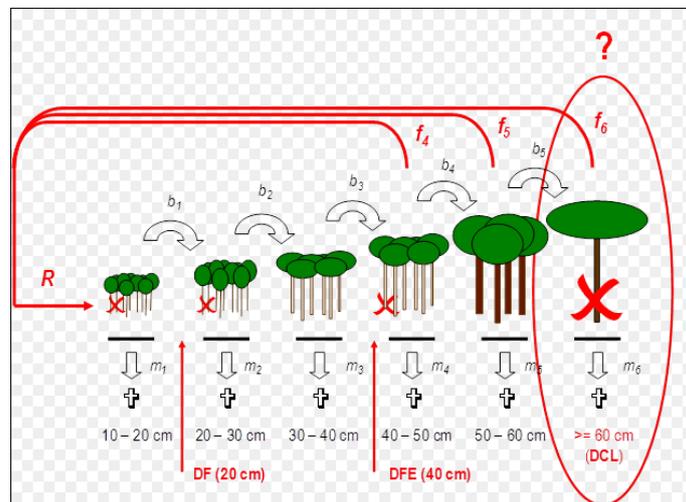


Figure 5 Processus d'évolution d'un peuplement [Dyn14c].

Pour atteindre ces objectifs des outils et des méthodes sont nécessaires, nous allons donc par la suite présenter ses éléments incontournables.

1.2. Les Méthodes et outils utilisés

1.2.1. Les outils

1.2.1.1. Le dispositif de M'Baïki

Le dispositif de **M'Baïki** est né de la volonté de l'Etat centrafricain de mieux gérer ses ressources forestières, et d'explorer les possibilités de reconstitution de la forêt dense humide du massif sud-ouest après un premier passage en exploitation forestière. Par ailleurs, il s'agit d'un dispositif de recherche permanent [AFM82]. Date de début du projet : 01/04/1982 et Date de fin du projet : 31/12/2022 [AFM82].

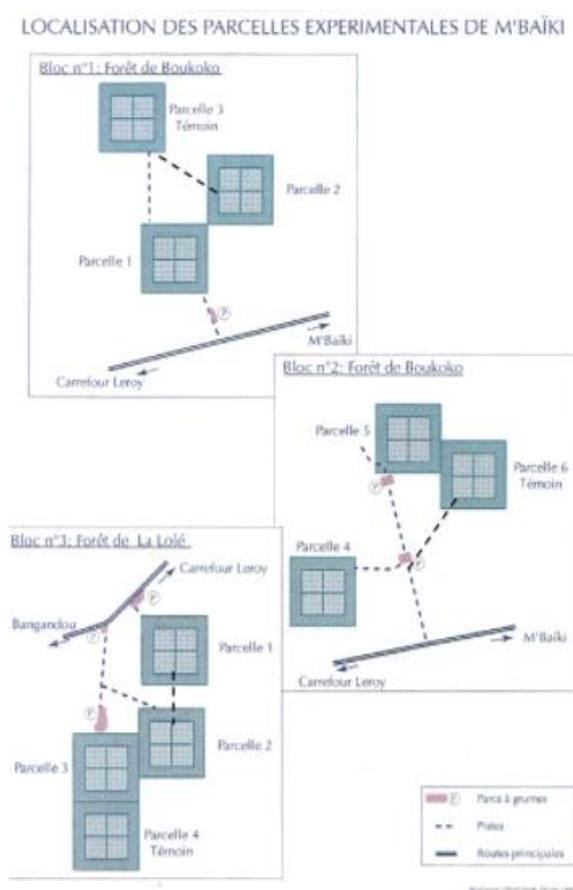


Figure 6 Localisation des parcelles expérimentales de M'Baïki

1.2.1.2. Les inventaires forestiers sur le dispositif M'Baïki

Les objectifs assignés à ce dispositif lors de sa délimitation en 1982 étaient d'apporter des réponses aux grandes questions suivantes :

- Quel sera le devenir des massifs forestiers centrafricains après un premier passage en exploitation?
- Comment les populations d'arbres exploitées reconstitueront-elles leurs effectifs ?

- Comment la forêt dense centrafricaine encore non exploitée se comporte-t-elle ?

Grâce à un suivi pratiquement continu depuis 1982, le dispositif a apporté la plupart des réponses à ces questions. Et Mbaïki est devenu un site de référence pour toute la région du bassin du Congo.

La longévité du site de Mbaïki et la qualité des suivis qui y ont été réalisés en font un dispositif totalement unique dans le bassin du Congo et l'un des très rares dispositifs dans le monde tropical permettant de répondre à ces nouvelles questions.

L'importance des données collectées à Mbaïki en font un outil très précieux sur lequel s'appuyer pour monter des projets d'étude des forêts dans le bassin du Congo. Ce dispositif constitue actuellement le socle du projet DynAffFor [AFM82].

1.2.1.3. Présentation du logiciel Utiliser pour la manipulation

Le stockage des données et leur manipulation est principalement réalisé dans ce cadre par le logiciel et langage R.

Présentation générale de R

R est un logiciel statistique créé par **Ross Ihaka** et **Robert Gentleman**. C'est à la fois un langage informatique et un environnement de travail. Les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont directement visualisés dans une fenêtre qui leur est propre [Lar15]. R est un langage orienté vers le traitement de données et l'analyse statistique dérivé du langage S. Il est développé depuis une vingtaine d'années par un groupe de volontaire de différents pays. C'est un logiciel libre² [MET17].

R présente plusieurs avantages à savoir :

- c'est un logiciel multiplateforme, qui fonctionne aussi bien sur des systèmes Linux, Mac Os ou Windows ;
- c'est un logiciel gratuit ;
- c'est un logiciel très puissant, dont les fonctionnalités de base peuvent être étendues à l'aide de plusieurs milliers d'extension ;
- c'est un logiciel dont le développement est très actif et dont la communauté d'utilisateurs ne cesse de s'élargir ;
- c'est un logiciel libre, développé par ses utilisateurs et modifiable par tout un chacun ;
- les possibilités de manipulation de données sous R sont en général largement supérieures à celles des autres logiciels ;
- c'est un logiciel avec d'excellentes capacités graphique et de nombreuses possibilités d'export ;
- R est de plus utilisé dans tous les secteurs scientifiques, y compris dans le domaine des analyses d'enquêtes et, plus généralement, des sciences sociales.

Sachant que rien n'est parfait, R possède également quelques inconvénients :

² Pour plus d'information sur ce qu'est un logiciel libre, voir : <http://www.gnu.org/philosophy/free-sw.fr.html>, publié sous licence GNU/GPL

- le logiciel, la documentation de référence et les principales ressources sont en anglais.
- Il est toutefois parfaitement possible d'utiliser R sans spécialement maîtriser cette langue ;
- comme R s'apparente davantage à un langage de programmation qu'à un logiciel. Proprement dit, la courbe d'apprentissage peut être un peu « raide », notamment Pour ceux n'ayant jamais programmé auparavant [MET17].

Philosophie de R

Quelques points particuliers dans le fonctionnement de R peuvent parfois dérouter les utilisateurs habitués à d'autres logiciels [Lar15] :

- Sous R, la production d'une analyse génère un grand nombre de résultats de toutes sortes avec un affichage réduit au minimum et c'est l'utilisateur qui demande à voir des résultats supplémentaires ou plus détaillés.
- Sous R, en général, on ne voit pas directement les données sur lesquelles on travaille ; on ne dispose pas en permanence d'une vue sur des données sous forme de tableau. Ceci peut être déroutant au début, mais on se rend vite compte qu'on n'a pas besoin de voir en permanence les données pour les analyser.
- Sous R, les résultats des analyses sont stockés dans les objets et sont dès lors manipulables. Inhabituel au début, ce fonctionnement permet en fait assez rapidement de gagner le temps dans la conduite des analyses [SGFVFABH98].

Notion d'extension ou « package »

L'installation par défaut du logiciel R contient le cœur du programme ainsi qu'un ensemble de fonction de base fournissant un grand nombre d'outils de traitement de données et d'analyse statistique [Lar15]. R étant un logiciel libre, il bénéficie d'une forte communauté d'utilisateurs qui peuvent librement contribuer au développement du logiciel en lui ajoutant des fonctionnalités supplémentaires. Ces contributions prennent la forme d'extension (packages) pouvant être installées par l'utilisateur et fournissant alors diverses fonctionnalités supplémentaires. Il existe un très grand nombre d'extension (plus de 6500 à ce jour), qui sont diffusées par un réseau baptisé CRAN.

La liste de toutes les extensions disponibles sur CRAN est disponible ici : <http://cran.r-project.org/web/packages/>.

Pour faciliter un peu le repérage des extensions, il existe un ensemble de regroupement schématiques (économétrie, finance, génétique, données, spatiales. . .) baptisés Task views³[MET17].

³ Pour plus d'information voir <http://cran.r-project.org/web/views>

1.2.2. Les méthodes

1.2.2.1. Brève Présentation des méthodes Monte Carlo

Description

Les techniques de simulation Monte Carlo (MC) sont utilisées pour simuler des systèmes déterministes ou stochastiques avec des paramètres ou des entrées incertaines ou stochastiques. Le nom a été proposé par les scientifiques du projet Manhattan lors de la deuxième guerre mondiale et fait allusion aux jeux de hasard pratiqués dans le casino de Monte Carlo dans la principauté de Monaco. Parmi les pionniers des méthodes MC nous retrouvons **E. Fermi, J. Neumann, S. Ulam, N. Metropolis.**

Nous utilisons aujourd'hui les méthodes MC pour simuler des phénomènes physiques complexes dans plusieurs domaines scientifiques et appliqués: radioactivité, physique des hautes énergies, réseaux, économétrie, logistique. La technique de simulation Monte Carlo s'appuie sur l'échantillonnage des distributions des quantités incertaines. Il peut être visualisé comme une boîte noire où entre un flux de nombres pseudo-aléatoires (c.-à-d. générés par l'ordinateur) et un flux dénombre sort; l'estimation de la quantité d'intérêt est obtenu en analysant là les données en sorties.

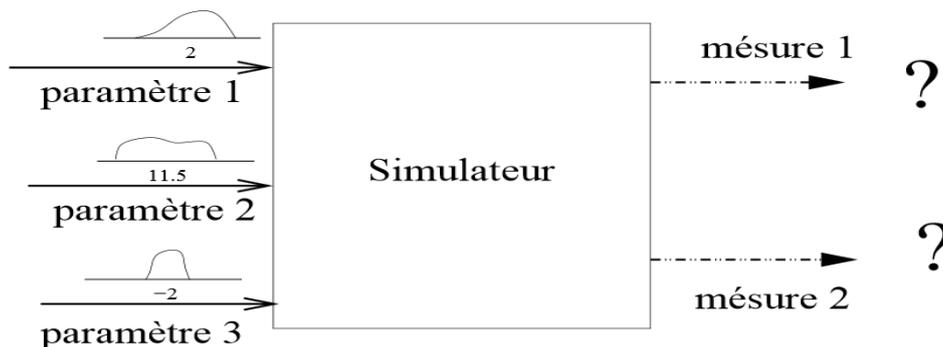


Figure 7 description de la méthode Monte Carlo [Bon09]

Composante d'un algorithme Monte Carlo

Description probabiliste: un modèle stochastique du problème. Générateur uniforme de nombres aléatoires: un générateur de nombres aléatoires uniformément distribués sur l'intervalle $[0,1]$. Loi d'échantillonnage: une technique pour échantillonner une distribution de probabilité générique. Simulateur: un simulateur déterministe qui renvoie l'output quand tous les paramètres sont connus. Collecteur des outputs: structure des données pour stocker tous les outputs de la simulation. Analyseur de l'output: ensemble de techniques statistiques qui permettent de tirer conclusions à partir des données générées par le simulateur. Estimateur d'erreur: ceci permet d'associer à chaque quantité estimée à partir de l'output une indication sur l'erreur ou sur la confiance (par exemple en fonction du nombre de répétitions).

1.2.2.2. Brève description sur les méthodes de calcul de certains indicateurs

Pendant le projet DynAffFor les experts en écologies et en modélisation de la dynamique forestière ont identifiés des indicateurs pertinents de la gestion durable des forêts. Ces indicateurs sont généralement recherchés par les gestionnaires des concessions forestières pour mieux raisonner la gestion de leurs écosystèmes. Certains de ces indicateurs nécessitent l'établissement de leur expression mathématique en fonction des données et des informations que nous disposons sur les arbres du peuplement forestier étudié [MET17].

Surface terrière

La surface terrière est une notion utilisée pour la gestion des peuplements forestiers irréguliers et pour leur régénération naturelle. Derrière cette définition se dissimule, en fait, un indicateur de la concurrence entre les arbres. La surface terrière n'a pas d'intérêt en soi : elle se compare à une valeur optimale qui diffère selon la composition en essences, l'âge des divers peuplements, l'objectif de production, etc. Si la densité d'un peuplement est supérieure à cette valeur, une intervention peut s'avérer pertinente pour favoriser la libre croissance des arbres [SiteA].

Définitions

Pour un arbre, c'est la surface de la section transversale de son tronc à 1,30 m de hauteur. Pour un peuplement, c'est la surface que représenteraient tous les troncs des arbres d'un hectare de la forêt que l'on aurait coupé à 1,30 mètre de hauteur [SiteA].

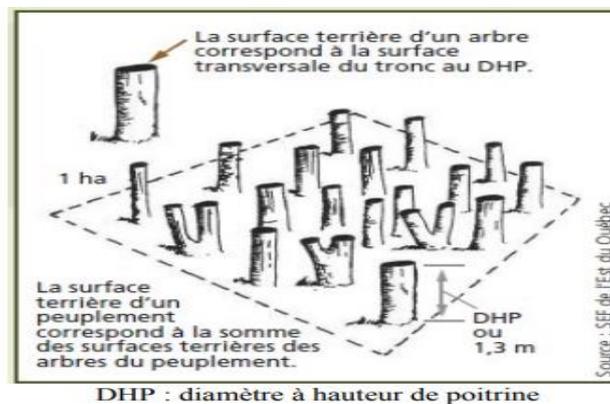


Figure 8 DHP: diamètre à hauteur de poitrine (siteA)

Cette grandeur un peu abstraite, notée G , dépend à la fois de la grosseur et du nombre d'arbres. La surface terrière est corrélée au couvert des arbres, ce qui permet de quantifier :

- le degré de compétition au sein du peuplement,
- les conditions d'éclairement au sol.

Elle représente donc très bien le volume et le stade du peuplement

- Un peuplement dense et âgé pourra avoir une surface terrière élevée 25 à 50 m²/Ha (chiffres valables pour des forêts tempérées)

- Un peuplement plus jeune ou plus clair aura une surface terrière faible : 5 à 15 m²/Ha. Cette donnée permet donc, à la fois, de savoir s'il faut intervenir dans le peuplement, mais aussi de calculer sommairement le volume sur pied à l'aide de la formule suivante :

On peut utiliser aussi certains tarifs de cubage (Tarifs de cubage «arbres» : les données de base sont prises à l'échelle de l'arbre (diamètre 1. 3, hauteur totale,... ; tarifs de cubage «peuplement» : les données de base sont des paramètres peuplements : **G,Hdom** par exemple)...Un Tarif de cubage «arbre» se présente sous la forme d'un tableau chiffré parfois aussi appelé « table de cubage», d'un graphique ou encore d'une équation qui fournit le volume d'un arbre en fonction d'une ou de plusieurs caractéristiques dendrométriques directement mesurables relatives à celui-ci [*SiteB*].

Mesure des surfaces terrières

Les données collectées sur le terrain, sont en réalité les circonférences des arbres mesurées à 1,3m. A partir de ces circonférences les diamètres et les surfaces terrières de chaque arbre sont déduites en appliquant les formules suivantes :

$$\text{Circonférence} = \pi * d$$

Nous mesurons unique le diamètre de l'arbre et par la formule suivante nous déterminons la surface terrière d'un arbre :

$$g_{1,3} = \frac{\pi}{4} d^2$$

La surface terrière G du peuplement est alors ensuite obtenue par sommation des surfaces Terrières g_i des N arbres à l'hectare le constituant, soit :

$$G = n_1 \cdot \frac{\pi}{4} \cdot d_1^2 + n_2 \cdot \frac{\pi}{4} \cdot d_2^2 + \dots + n_j \cdot \frac{\pi}{4} \cdot d_j^2 = \frac{\pi}{4} \cdot \sum n_i d_i^2 = \sum n_i g_i$$

L'unité de mesure de G est le m² /ha. Comme les circonférences et les diamètres sont mesurés en cm, il ne faut pas oublier de les convertir en m avant de calculer la surface terrière.

Connaissant la surface terrière d'un peuplement (G), on peut en déduire la surface terrière moyenne et le diamètre qui lui correspond.

Sur la base de G et de N, on peut calculer la surface terrière moyenne [MET17].

$$\bar{g} = G/N.$$

Volume total de bois

D'une manière générale la détermination du volume peut se faire de deux façons différentes : les méthodes directes et les méthodes indirectes [VIN11]. Pour les méthodes directes la détermination peut se faire aussi de deux façon différentes selon que l'arbre est abattu (également appelé méthode destructive) ou sur pied. Les méthodes d'estimation du volume sont généralement appelées tarifs de cubage.

Volume exploitable

Correspond au volume d'arbre du peuplement forestier dont le diamètre est au supérieur au $DME^{2\wedge}$.

La Biomasse

Jusque récemment, la dendrométrie s'orientait principalement vers l'estimation du volume car celui-ci servait de base pour les ventes de bois. Cependant, ces dernières années, la biomasse a connu un important regain d'intérêt. Celui-ci s'explique par le fait que la masse de bois est souvent préférée dans certains secteurs tels que l'industrie papetière et le bois énergie. Elle permet de mieux comprendre le fonctionnement des écosystèmes forestiers et comparer leurs productivités. De plus la biomasse est à la base des calculs de stocks de carbone en forêt et donc en lien direct avec les études sur le changement climatique

Structure diamétrique

La structure diamétrique est la répartition des tiges par classes de diamètres, elle peut s'établir en prenant en compte tous les individus et toutes espèces confondues. Elle est porteuse d'information sur la stabilité de toutes les espèces confondues. Elle est peut être également s'établir par espèce et l'on parle de structure spécifique.

Stock d'arbres

$N = \sum n_i$ Avec n_i l'effectif du groupe d'espèce dans la classe de diamètre i .

Stock exploitable

Correspond au stock d'arbre du peuplement forestier dont le diamètre est au supérieur au DME^4

Récapitulatif

Nous venons de voir les concepts inhérents à notre problématique, nous avons aussi présenté dans un premier temps un état des lieux sur le simulateur existant qui constitue la base de notre travail, en présentant son principe de fonctionnement et le modèle statistique sur lequel il est fondé, puis les outils utilisés dans le cadre de son implémentation. Et dans un second temps nous avons présenté les méthodes de calculs de certains éléments d'évaluation de la gestion durable des forêts appelés les indicateurs forestiers.

⁴ Diamètre minimum d'exploitation

Chapitre 2 : Mise en œuvre

Introduction

Dans le but de suivre l'évolution d'une forêt, il est nécessaire de modéliser les trois composantes de sa dynamique : la régénération, la croissance et la mortalité. Sachant qu'une forêt est un peuplement d'arbres hétérogènes rendant sa description complexe à cause de sa grande diversité spécifique, nous nous proposons de regrouper les espèces d'arbres en groupes ayant un comportement similaire suivant les classes diamétriques et les trois processus. Ce qui nous permettra de partitionner le peuplement forestier, afin de décrire l'évolution de chaque espèce dans chaque classe de diamètre et ainsi prédire la dynamique globale de la forêt. Notre travail est une contribution pour étendre le champ d'action du logiciel **DafSim** basé sur le modèle **DafMod**. Ceci en tenant compte de la possibilité de l'irrégularité des intervalles de temps entre les inventaires de données forestiers et par la suite proposé une intégration au logiciel **DafSim** par la modélisation d'un algorithme propre à notre problème.

2.1. Définition globale du problème

Pour permettre au Simulateur du peuplement forestier **DafSim** de bien fonctionner et de fournir de bons résultats, il est impératif de fournir des inventaires sur le peuplement cible à des intervalles de temps réguliers et si possible rapproché (idéalement tous les ans).

Mais pour une raison ou une autre à savoir, des guerres civiles, des intempéries et mêmes pour des raisons politiques etc..., il peut arriver d'être dans l'impossibilité totale de lancer une opération d'une si grande envergure à savoir un inventaire forestier. L'inventaire forestier vise à connaître la ressource en bois disponible. C'est le cas du modèle suivant :

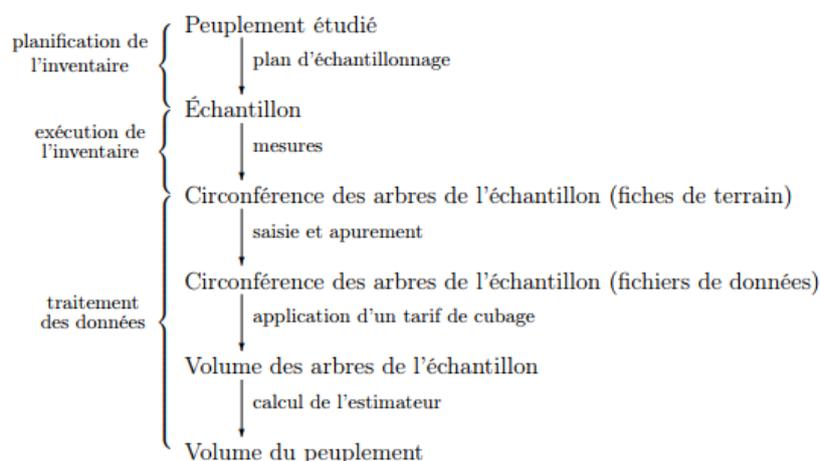


Figure 9 Chaîne des opérations de l'inventaire forestier ([CMUC17])

Soucieux de pallier à ce problème, il sera question d'améliorer la modélisation proposée par **BERTRAND TEUGUIA** par une modélisation adaptée au contexte et de produire un algorithme pour le cas général des inventaires de données collectées à des intervalles de temps quelconque.

2.1.1. Modélisation pour les intervalles de Temps irréguliers entre les inventaires

La modélisation est le procédé par lequel nous utilisons des expressions mathématiques pour décrire une situation quantitative réelle. Modéliser consiste à écrire en notation mathématique, ce qui est exprimé de façon littéraire en faisant intervenir des variables au besoin.

Dans cette partie nous entendons par inventaires à intervalles de temps irréguliers, le fait que le temps qui sépare deux inventaires consécutifs n'est pas toujours le même.

2.1.1.1. Modèle du simulateur

Hypothèse pour le simulateur

Notre modélisation s'appuiera sur un certains nombres d'hypothèses à savoir :

- Hypothèse de Markov : l'évolution du peuplement entre 2 temps t et $t+1$ ne dépend que de son état au temps t .
- Hypothèse indépendance : les arbres sont indépendants.
- Hypothèse de stationnarité : le passage d'un arbre d'un état vers un autre dépend du temps.
- Hypothèse d'USHER : A un temps t un arbre peut soit mourir, soit passer à la classe supérieure soit rester dans la même classe. (confert These Dakis) [DKPD11].

Notion de classe de taille

Ici, nous modéliserons la dynamique d'un arbre au travers de différentes classes de tailles tout en gardant à l'esprit ici qu'une taille en fait est égale à : DHP (Diamètre à Hauteur de Poitrine)

Soit par exemple un ensemble de 5 classes :

- Classe 1 :] 20,40 [
- Classe 2 :] 30,40 [
- Classe 3 :] 40,50 [
- Classe 4 : [50,60]

Associées à 4 états 1, 2, 3, 4 par exemple soit le tableau suivant :

Tableau 1 classe de diamètre

Classe] 20,40 [] 30,40 [] 40,50 [[50,60]
Etat	1	2	3	4

L'adjonction de l'état « 0 » à ces états représente en fait la mort d'un individu, d'un arbre. Dans ce genre de modèle le passage d'une classe à une autre signifie alors un changement d'état. A l'état 4 l'individu reste alors dans cet état ou alors peut disparaître [VRFM2010].

En résumé : chaque individu dans notre simulateur appartiendra à une classe représenté ci-dessus. Il peut passer d'une classe à la suivante ceci en fonction de sa croissance ou de sa taille [UMB07].

Notion de vraisemblance

Vraisemblance Caractère de ce qui est vraisemblable, a l'apparence de la vérité (le petit Larousse).

Elle est une des idées de base de la statistique. Elle donne un cadre général et très puissant pour traiter toutes sortes d'applications, en particulier pour :

- La détermination de la ou des variables de θ pour lesquelles la fonction de vraisemblance sera maximale.
- construire des tests puissants [SiteC].

Dans le cadre de notre travail la vraisemblance mesurera une adéquation entre la distribution observée (les inventaires de données) sur un échantillon aléatoire (les parcelles de M'Baïki) et les lois de probabilité qui sont supposée décrire la réalité sur la population (la croissance, la mort et la régénération dans le bassin du Congo) dont l'échantillon est issu, Comment ?

Par hypothèse, nous connaissons la fonction de densité de la loi de probabilité théorique la mieux adaptée à la population. Cependant, c'est une COQUILLE VIDE puisque nous ignorons ses paramètres (espérance, variance). N'ayant pas le choix, nous estimons ceux-ci à partir de Statistiques d'échantillon.

Il existe alors plusieurs techniques pour trouver l'estimateur de plus pertinent, et donc l'une d'elle est l'utilisation de la **Vraisemblance** [SiteD].

Son expression est généralement donnée par :

$$P(x_1, x_2, \dots, x_n; \theta) \quad (2.18)$$

Le grand intérêt de cette fonction est de permettre la détermination de la ou des valeurs de θ pour lesquelles la fonction de vraisemblance sera **maximale** [SiteD].

2.1.1.2. Modélisation de la croissance

Processus de croissance

D'une manière générale, la croissance d'un individu peut se définir comme la variation dans le temps des variables qui permettent sa description. Dans le cas d'un arbre, ces variables doivent tenir compte de:

- l'absorption de l'eau
- l'absorption d'éléments minéraux
- la nutrition azotée
- la photosynthèse

De manière plus spécifique, **la croissance** : est le fait pour un arbre de passer de sa classe de diamètre à la classe de diamètre directement au-dessus.

Modèle pour la croissance

Ici il s'agit pour nous, d'exprimer de manière formelle la variation temporelle des grandeurs descriptives de l'arbre en fonction des variables que l'on peut classer parmi les catégories suivantes :

- Les grandeurs descriptives de l'arbre elles-mêmes(X), typiquement son diamètre à 1, 30m (D), sa hauteur, etc.
- Les variables de compétition(C), appelées aussi indice de compétition
- Les variables de site décrivant l'effet du milieu(S)

Remarque : la plupart des auteurs s'accordent à reconnaître les variables de dimension comme les meilleurs prédicateurs de la croissance individuelle. La dimension traduisant à la fois les performances passées de l'individu ainsi que son statut social dans le peuplement. Les autres variables sont quasiment insignifiantes.

Il est important de noter que, Le plus souvent, l'accroissement de X à l'instant t est prédit à partir de la valeur de X à cet instant :

$$\frac{dy}{dx} = f(X, C, S) \quad (2.1)$$

Où f est appelé fonction de croissance. Et pour la cohérence du système, on ajoute des équations dites d'état qui relient les variables descriptives entre elles :

$$X_i = g(X_j), \quad i \neq j \quad (2.2)$$

C'est donc la connaissance des formes des fonctions f de 2 et g de 3 qui fait le cœur de la modélisation de la croissance.

En pratique

Nous allons répéter des observations sur différentes parcelles de la forêt sachant que ces parcelles contiennent plusieurs arbres d'espèces différentes nous aurons le tableau suivant :

Tableau 2 Tableau de données pour les croissances, $i= 1 \dots N$ ([TEM2017])

Y		X							
g	n	Esp	$X_1^{(init)}$	\dots	$X_L^{(init)}$	$X_1^{(end)}$	\dots	$X_L^{(end)}$	d
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$g_{s,i}$	$n_{s,i}$	s	$x_{s,i,1}^{(init)}$	\dots	$x_{s,i,L}^{(init)}$	$x_{s,i,1}^{(end)}$	\dots	$x_{s,i,L}^{(end)}$	$d_{s,i}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

- g désigne le nombre d'arbres qui ont grandi ;
- n désigne le nombre d'arbre observés ;
- d la période qui s'est écoulé jusqu'à l'année i
- 'Esp' désigne l'espèce des arbres observés;
- $X_1^{init} \dots X_L^{init}$ désigne les variables explicatives pour une parcelle à une année initiale, chacune donnant la surface terrière ou le nombre d'arbre de la classe diamétrique de son indice i ;
- $X_1^{end} \dots X_L^{end}$ désigne les variables explicatives pour la même parcelle mais après une période 'd'exprimée en années après l'année initiale d'observation, chacune donnant aussi la surface terrière de la classe diamétrique de son indice.

Remarque : plus généralement nous utiliserons la notation en exposant $(.)^{(t)}$ pour indiquer les paramètres et les variables entre (init) et (end) et chaque ligne représente les parcelles où plusieurs observations sur les arbres d'une espèce ont été faites.

Puisque les données de chaque ligne représentent les observations sur plusieurs arbres avec une période 'd' (en années), nous allons procéder par la modélisation pour un seul arbre afin de parvenir à la modélisation pour une espèce.

En effet, **La croissance** : est le fait pour un arbre de passer de sa classe de diamètre à la classe de diamètre directement au-dessus.

La variable aléatoire observée ne peut prendre que deux valeurs que l'on identifie par les labels de $\{0, 1\}$. C'est donc d'après la valeur reçue, que nous allons ainsi pouvoir dire si l'arbre change de classe de diamètre (l'évènement : a grandi) ou alors reste dans la même classe (l'évènement : il n'a pas grandi).

Nous souhaitons estimer la probabilité inconnue qu'un individu "a" grandi ou pas. Ainsi donc pour **modéliser la croissance nous avons choisi le modèle de régression logistique** car c'est un modèle **binomial**, qui nous permet d'expliquer exclusivement les variables binaires. Ici la régression linéaire fait face à un problème de taille, en effet soit l'équation :

$$y = \theta X + \epsilon \quad (2.3)$$

« ϵ » représente le bruit qui d'après la définition d'un modèle de régression linéaire suit une loi Normale ; hors dans notre cas Y suit une loi de Bernoulli ce qui ne permet pas de définir une relation valide entre les variables ($Y \in [0, 1]$ et $X\beta + \epsilon \in [0, 1]$) , **le modèle de régression logistique** nous propose dans ce cas de modifier l'expression de la variable à expliquer en appliquant une fonction appelée fonction **logit (la transformation logit)** à l'espérance de la variable aléatoire de Y [CMUC17] :

$$\text{logit}(y) = \ln\left(\frac{y}{1-y}\right) \in \mathbb{R} \text{ et } y \notin]0, 1[\quad (2.4)$$

Pour une régression à plusieurs variables explicatives définissons les expressions suivantes :

- $P(Y)$ est la probabilité que Y arrive (son espérance car Y suit une loi de Bernoulli)
- $\theta = (\theta_0 \theta_1 \dots \dots \dots \theta_n)$ Les coefficients représentent la combinaison linéaire du prédicteur et de la constante.
- $X_1 \dots \dots X_n$ Les variables explicatives

Alors en utilisant l'inverse de la fonction LOGIT, nous obtenons la probabilité de succès pour un arbre en fonction de certaines variables qui expliquent son comportement.

$$P(Y) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \dots + \theta_L X_L)}} \quad (2.5.1)$$

En posant $x_{s,j}^{(t)} = (1, x_{s,j,1} \dots \dots \dots x_{s,j,L})$ le vecteur de données des variables explicatives de l'individu à l'instant t (les arbres de l'espèce s d'une parcelle) à la ligne j, et en considérant la variable aléatoire réponse de cette ligne $y_{s,j}$, comme celle donnant la croissance sur tous les $n_{s,j}$ d'arbres de la ligne i, $y_{s,j}$ est la répétition de la loi de Bernoulli, donc une loi Binomiale $\beta(n_{s,j}, p_{s,j})$.

Cette forme de modèle logistique est sa forme générale, on parle de modèle logistique à données groupées ou répétées. L'observation de la variable $y_{s,j}$ est pour l'espèce s vue comme le rapport $\frac{g_{s,j}}{n_{s,j}}$. [CMUC17]

$$p\left(y_{s,j} = \frac{g_{s,j}}{n_{s,j}}\right) = \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}} \quad j \in \{1, 2 \dots L\} \text{ et } t \in [1, d_{s,i} - 1] \quad (2.6.2)$$

Ce modèle de régression est celui d'une espèce. Les classes de diamètre étant généralement de plage 10 cm, même en cinq ans ou plus il n'est pas réaliste d'envisager qu'un arbre ait changé deux fois de suite sa classe de diamètre..

Le modèle à intervalles de temps réguliers (d'un an) entre les inventaires représente pour nous le modèle idéale. C'est par le modèle régulier (**confère [TEM2017]**) que nous définissons la

croissance pour un arbre en fonction des surfaces terrières, hors le problème ici c'est que nous ignorons les paramètres issus d'une régularité car il peut s'écouler plus d'une année avant un nouveau inventaire. Nous allons donc exprimer ses paramètres inconnus par rapport aux paramètres d'inventaires réguliers. Ainsi, il nous faut trouver le lien entre la probabilité de croissance d'inventaires distants de « d » années en fonction des probabilités d'inventaires des années non observées. On a le schéma suivant :

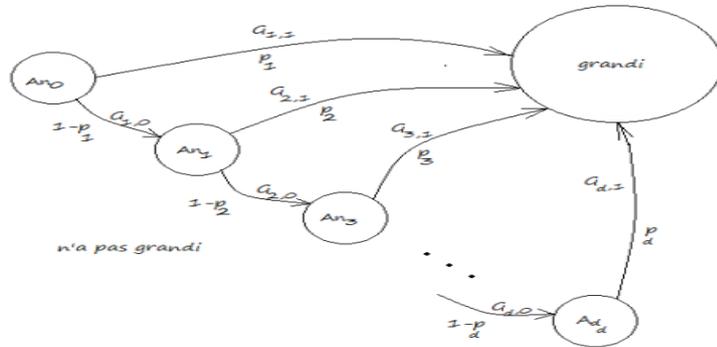


Figure 10 Automate de croissance ou de morali te pour d ann es (TEM2017)

Dans cette figure $G_{i,1}$: est l' v nement : 'grandi l'ann e i de probabilit  p_i et $G_{i,0}$: est l' v nement : « n'a pas grandi l'ann e i » ainsi la formule du calcul de la probabilit  de grandir sur ' n ' ann es est celle d'une r union d' l ments, et on peut la calcul e par la formule de Poincar  :

Puisque les inventaires ne sont pas effectu s chaque ann e, posons :

- J est l'ensemble des ann es o  on fait les inventaires.
- $p_{s,k}$: D signe pour l'esp ce s la Probabilit  qu'un arbre grandit pendant l'ann e k
- $P_{d,s,j}$: D signe pour l'esp ce s la probabilit  qu'un arbre grandit $d_{s,j}$ ans jusqu'  l'ann e J .

$$\begin{aligned}
 P_{d,s,j} &:= \sum_{k=1}^{d_{s,j}} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq d} p_{i_1} p_{i_2} \dots p_{i_k} \\
 &= \sum_{i=1}^{d_{s,j}} p_i \prod_{K < i} (1 - p_K) \\
 &= 1 - \prod_{K=1}^{d_{s,j}} (1 - p_K) \tag{2.7}
 \end{aligned}$$

Par exemple pour $d_j = 2$ on a :

$$P_2 = p_1 + p_2 - p_1 p_2 = 1 - (1 - p_1)(1 - p_2)$$

Soit donc :

$\mathbb{P}(Y_{s,j} = g_{s,j})$: Probabilité que $d_{s,j}$ années jusqu'à l'année 'j', il y'ait g_i arbres qui aient grandi sur les $n_{s,j}$ arbres observés. Puisque la variable Y suit la loi **Binomial**.

$$\mathbb{P}(Y_{s,j} = g_{s,j}) = \binom{g_{s,j}}{n_{s,j}} p_{d_{s,j}}^{g_{s,j}} (1 - p_{d_{s,j}})^{(n_{s,j} - g_{s,j})} \quad (2.8)$$

Sachant que $p_{d_{s,j}}$ correspond à une probabilité de croissance sur $d_{s,j}$ années, en utilisant(2.7), on a :

$$\mathbb{P}(Y_{s,j} = g_{s,j}) = \binom{g_{s,j}}{n_{s,j}} \left(1 - \prod_{K=1}^{d_{s,j}} (1 - p_K)\right) \left(1 - \left(1 - \prod_{K=1}^{d_{s,j}} (1 - p_K)\right)\right)^{(n_{s,j} - g_{s,j})}$$

$$\mathbb{P}(Y_{s,j} = g_{s,j}) = \binom{g_{s,j}}{n_{s,j}} \left(1 - \prod_{K=1}^{d_{s,j}} (1 - p_K)\right)^{g_{s,j}} \left(\prod_{K=1}^{d_{s,j}} (1 - p_K)\right)^{(n_{s,j} - g_{s,j})}$$

Sachant que $p_{s,K} = \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}} \quad j \in \{1, 2 \dots L\}$

$$\mathbb{P}(Y_{s,j} = g_{s,j}) = \binom{g_{s,j}}{n_{s,j}} \left(1 - \prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{g_{s,j}} \left(\prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{(n_{s,j} - g_{s,j})} \quad (2.9)$$

- La vraisemblance sur les J années d'observations est donnée par :

$$L(X; \theta) = \prod_{j \in J} \mathbf{P}(Y_{s,j} = g_{s,j})$$

$$L(X; \theta) = \prod_{j \in J} \binom{g_{s,j}}{n_{s,j}} \left(1 - \prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{g_{s,j}} \left(\prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{(n_{s,j} - g_{s,j})} \quad (2.9)$$

Le log de la vraisemblance est donnée par :

$$L(X; \theta) = \sum_{j \in J} \ln \left(\binom{g_{s,j}}{n_{s,j}} \left(1 - \prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{g_{s,j}} \left(\prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{(n_{s,j} - g_{s,j})} \right)$$

Le terme $\binom{g_{s,j}}{n_{s,j}}$ n'influencera pas lors de la maximisation de la vraisemblance, donc nous allons le négligé.

D'où :

$$\begin{aligned} L(X; \theta) &= \sum_{j \in J} \ln \left(\left(1 - \prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{g_{s,j}} \left(\prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right)^{(n_{s,j} - g_{s,j})} \right) \\ L(X; \theta) &= \sum_{j \in J} g_{s,j} * \ln \left(1 - \prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right) + \sum_{K=1}^{d_{s,j}} (n_{s,j} - g_{s,j}) * \ln \left(\prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right) \\ &= \sum_{j \in J} g_{s,j} * \ln \left(1 - \prod_{K=1}^{d_{s,j}} \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right) + (n_{s,j} - g_{s,j}) * \sum_{K=1}^{d_{s,j}} \left(\ln \left(1 - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}}\right)\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{j \in J} g_{s,j} * \ln \left(\mathbf{1} - \prod_{K=1}^{d_{s,j}} \left(\mathbf{1} - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}} \right) \right) + (n_{s,j} - g_{s,j}) * \sum_{K=1}^{d_{s,j}} \left(\ln \left(\frac{e^{-(\theta * x_{s,i}^{(t)})}}{1 + e^{-(\theta * x_{s,i}^{(t)})}} \right) \right) \\
&= \sum_{j \in J} g_{s,j} * \ln \left(\mathbf{1} - \prod_{K=1}^{d_{s,j}} \left(\mathbf{1} - \frac{1}{1 + e^{-(\theta * x_{s,i}^{(t)})}} \right) \right) + (n_{s,j} - g_{s,j}) * \sum_{K=1}^{d_{s,j}} \left(\ln \left(e^{-(\theta * x_{s,i}^{(t)})} \right) - \ln \left(1 + e^{-(\theta * x_{s,i}^{(t)})} \right) \right) \\
&= \sum_{j \in J} g_{s,j} * \ln \left(\mathbf{1} - \prod_{K=1}^{d_{s,j}} \left(\frac{e^{-(\theta * x_{s,i}^{(t)})}}{1 + e^{-(\theta * x_{s,i}^{(t)})}} \right) \right) + (n_{s,j} - g_{s,j}) * \sum_{K=1}^{d_{s,j}} \left(-(\theta * x_{s,i}^{(t)}) - \ln \left(1 + e^{-(\theta * x_{s,i}^{(t)})} \right) \right) \quad (2.10)
\end{aligned}$$

2.1.1.3. Modélisation du recrutement

Processus de recrutement

Le Recrutement, est le fait qu'un arbre juvénile intègre la première classe de diamètre et devient pris en compte dans le peuplement.

On s'intéresse à un grand nombre d'événements indépendant. Evénement qui survient au hasard. On observe qu'ils se produisent λ fois en moyenne durant un intervalle de temps donnée.

Nous voulons ici compter le nombre d'espèces d'arbre qui poussent pour chaque espèce suivant une période donnée en fonction de certaines variables explicatives. Nous avons donc choisie La régression de Poisson qui est un modèle utilisé pour modéliser les données de comptage.

Modèle de recrutement

Soit le tableau représentatif suivant, il représente des observations répétées sur différentes parcelles de la forêt sachant que ces parcelles contiennent plusieurs arbres d'espèces différentes nous aurons le tableau suivant :

Tableau 3 pour le recrutement $i=1 \dots N$ ([TEM2017])

Y	X							
r	Esp	$X_1^{(init)}$	\dots	$X_L^{(init)}$	$X_1^{(end)}$	\dots	$X_L^{(end)}$	d
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$r_{s,i}$	s	$x_{s,i,1}^{(init)}$	\dots	$x_{s,i,L}^{(init)}$	$x_{s,i,1}^{(end)}$	\dots	$x_{s,i,L}^{(end)}$	$d_{s,i}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

- \mathbf{r} désigne le nombre d'arbres recrutement ;
- \mathbf{n} désigne le nombre d'arbres observés ;
- \mathbf{d} la période qui s'est écoulé jusqu'à l'année i

- $r_{s,i}$ désigne pour l'espèce s , le nombre moyen de recrût à la ligne i .
- 'Esp' désigne l'espèce des arbres observés.
- $X_i^{init} \dots X_L^{init}$ désigne les variables explicatives pour une parcelle à une année initiale, chacune donnant la surface terrière ou le nombre d'arbre de la classe diamétrique de son indice i .
- $X_i^{end} \dots X_L^{end}$ désigne les variables explicatives pour la même parcelle mais après une période d exprimée en années après l'année initiale d'observation, chacune donnant aussi la surface terrière de la classe diamétriques de son indice.
-

Remarque : Chaque ligne représente les parcelles où plusieurs observations sur les arbres d'une espèce ont été faites.

Comme dans le cas précédant, on cherche à exprimer le paramètre de la loi de Poisson après une période d exprimée en années, en fonction des paramètres des années intermédiaires. Nous partons du fait que le paramètre de la loi de Poisson est aussi son espérance et puisque nous sommes dans une situation de somme de telles lois (qui est encore une loi de Poisson), on déduit alors que le paramètre λ correspondant vérifie :

$$\begin{aligned}
 \lambda &= E(P(\lambda)) \\
 &= E(P(\lambda_1) + P(\lambda_2) + \dots + P(\lambda_d)) \\
 &= E(P(\lambda_1)) + E(P(\lambda_2)) + \dots + E(P(\lambda_p)) \\
 \lambda &= \lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_d
 \end{aligned}
 \tag{2.11}$$

Si on note $y_{s,i}$ la variable aléatoire réponse de la ligne i , on sait d'après [TEM2017] (page 35 et 51) dans le cas régulier nous savons que:

$$\lambda_{s,i} = \exp(\theta * x_{s,i}) \text{ Avec } i \in \{1, \dots, L\} \tag{2.11}$$

Et que pour l'espèce s on a :

$$P(y_{s,1}, \dots, y_{s,j}, \dots, y_{s,n_s} = r_{s,1}, \dots, r_{s,j}, \dots, r_{s,n_s}) = \prod_{j=1}^{n_s} \frac{\lambda_{s,j}^{r_{s,j}}}{r_{s,j}!} \exp(-\lambda_{s,i})
 \tag{2.12}$$

Par (2.11) On a la relation suivante dans le cas irrégulier :

$$\lambda_{s,j} = \sum_{j=1}^{d_{s,j}} \lambda_{s,j}^{(t)} = \sum_{j=1}^{d_{s,j}} \exp(\theta * x_{s,j}^{(t)})$$

- La vraisemblance sur les d années d'observations est donnée par :

$$L(X; \theta) = \prod_{j \in J} \left(\frac{\sum_{t=1}^{d_{s,j}} \lambda_{s,j}^{(t)}}{r_{s,j}!} \right) * \exp(-\sum_{j=1}^{d_{s,j}} \lambda_{s,j}^{(t)})$$

$$L(X; \theta) = \prod_{j \in J} \left(\left(\frac{\sum_{j=1}^{d_{s,j}} \exp(\theta * x_{s,j}^{(t)})}{r_{s,j}!} \right) * \exp(- \sum_{j=1}^{d_{s,j}} \exp(\theta * x_{s,j}^{(t)})) \right)$$

(2.13)

On déduit la log-vraisemblance suivante :

$$L(X; \theta) = \sum_{j \in J} \left(\left(\frac{\sum_{j=1}^{d_{s,j}} \exp(\theta * x_{s,j}^{(t)})}{r_{s,j}!} \right) * \exp(- \sum_{j=1}^{d_{s,j}} \exp(\theta * x_{s,j}^{(t)})) \right) \quad (2.14)$$

2.1.1.4. Modélisation de la mortalité

Processus de mortalité

Dans le cadre de notre travail La mortalité accidentelle n'est pas en générale considérée dans la modélisation. La mortalité régulière se modélise quant à elle soit de façon déterministe soit de façon stochastique, ce dernier cas est le plus souvent observé. Nous allons alors ne considérer que la probabilité de mourir pour l'arbre est constante, variable ou dépend des caractéristiques de l'arbre et/ou du peuplement, la plus recommandée étant cette dernière. Les variables les plus recommandées servant à l'estimation de la probabilité de mourir d'un arbre sont :

- Le diamètre et/ou la surface terrière (surface à 1,30 m du sol) du sujet ;
- Un indice de compétition ;
- Une variable décrivant l'accroissement durant la période précédente, en diamètre ou en surface terrière. Le modèle préconiser pour relier ces variables aux probabilités (ou taux) de mortalité des arbres est le modèle logistique.

Modèle de mortalité

La mortalité et la croissance dans notre contexte de représentation du peuplement en classe de diamètre, se modélisent de façon tout à fait similaire, la croissance a un comportement d'état absorbant car dans une classe, on la considère similaire à la mort. La mort naturelle d'un arbre est un évènement rare. C'est par le modèle logit-binomial défini pour la croissance que nous définissons la mortalité pour un arbre en fonction des surfaces terrières.

Remarque : la cible visée dans ses modèles est la détermination des coefficients réels inconnus $\theta = (\theta_0 \theta_1 \dots \theta_n$ caractérisant le lien existant entre la variable réponse Y et les variables explicatives $X_1 \dots X_n$; on souhaite les estimer à l'aide des données des inventaires.

2.1.2. Interpolation des données de variables explicatives

Problématique

Les inventaires de données peuvent être réalisés une année (année initiale) et ensuite après une période d exprimée en années supérieure à 1, nous n'avons pas dans ce cas de figure les données des années où les inventaires n'ont pas été faits ; **hors** nous devons connaître ces informations pour être en mesure de calculer la vraisemblance

Cette partie est essentiellement inspirée du document [TEM2017] (page 6)

Présentation de la méthode d'interpolation (par noyau)

L'objectif de trouver une relation entre les inventaires d'une même parcelle distante d'une distance d (en années). Pour une donnée absente, l'idée d'approximation de sa valeur est de pondérer les données présentes qui l'entourent en attribuant plus d'importance à la donnée qui lui est plus proche [ASG09]. Nous entendons par données, le vecteur X de taille $L + 1$ dont les composantes de 2 à $L + 1$ représentent des valeurs de surface terrière.

Soit $x_{s,i}^{(t)}$ la t^e donnée manquante avec $t \in [1, d_{s,i} - 1]$ et $x_{s,i}^{(d_{s,i})} = x_{s,i}^{(end)}$ et $x_{s,i}^{(0)} = x_{s,i}^{(init)}$.

On se servira des données observées $x_{s,i}^{(end)}$ et $x_{s,i}^{(init)}$ pour approximer la valeur de $x_{s,i}^{(t)}$ $t \in [1, d_{s,i} - 1]$, il s'agit de trouver une relation de la forme :

$$x_{s,i}^{(t)} = \omega_{init} x_{s,i}^{(init)} + \omega_{end} x_{s,i}^{(end)} \quad (2.15)$$

Où les poids ω dépendent de l'inventaire manquant et la durée entre les inventaires observés. Il s'agit d'écrire la donnée $x_{s,i}^{(t)}$ comme barycentre des données $x_{s,i}^{(end)}$ et $x_{s,i}^{(init)}$ de telle sorte que les poids soient inversement proportionnelles à leurs distances à $x_{s,i}^{(t)}$. On choisit les poids de la manière suivante :

$$\omega(t) = \frac{\mathcal{P}(|t-T|)}{\mathcal{P}(t)+\mathcal{P}(|t-T|)} \quad (2.16)$$

$T \in \{(init), (end)\}$, \mathcal{P} , est une fonction positive et décroissante des distances à t . Nous choisissons $\mathcal{P}(a) = d_{s,i} - a$. Ce choix nous permet de rejoindre la méthode d'interpolation linéaire. En fonction du comportement des données, on peut être amené à le définir autrement.

Formule de complétion des données

Supposons que $d_{a,i} = 3$ et que (init) correspond aux données de 1989 et (end) 1992, en appliquant la (2.15) à 1989 ($t = 1$) avec $\mathcal{P}(a) = d_{a,i} - a$, on obtient :

$$\begin{cases} \mathcal{P}(|(init) - t|) = \mathcal{P}(|0 - t|) = 3 - 1 = 2 \\ \mathcal{P}(|(end) - t|) = \mathcal{P}(|d_{s,i} - 1|) = 3 - (3 - 1) = 1 \\ \mathcal{P}(t) = 3 - 1 = 2 \end{cases}$$

D'où :

$$x_{s,i}^{(1)} = \frac{3 - 1}{(3 - 1) + (3 - 2)} x_{s,i}^{(init)} + \frac{3 - 2}{(3 - 1) + (3 - 2)} x_{s,i}^{(end)}$$

$$x_{s,i}^{(1)} = \frac{2}{3} x_{s,i}^{(init)} + \frac{1}{3} x_{s,i}^{(end)}$$

On peut remarquer que ce résultat est très intuitif. En effet, le poids pour un point non observé est défini comme étant le rapport entre l'écart complémentaire au point non observé et la valeur manquante, à la distance séparant les deux points observés (init) et end).

Soit donc la formule de complétion des données suivante :

$$x^{(t)} = \frac{d_{s,i} - t}{d_{s,i}} x_{s,i}^{(init)} + \frac{t}{d_{s,i}} x_{s,i}^{(end)} \quad (2.17)$$

Nous avons présenté dans cette partie une méthode générale d'approximation des valeurs manquantes. Nous avons pu constater que celle-ci dépend de la définition des poids de pondération des observations. Pour notre cas nous avons choisi comme poids le rapport entre l'écart complémentaire du point observé et la valeur manquante sur la distance séparant les deux points observés (init) et (end), sur cette dernière. Par ailleurs, il existe d'autres moyens de définition des poids qui peuvent être liés directement à l'expérience, celles-ci reposent entièrement sur la détermination de la fonction \mathcal{P} qui permet d'avoir les poids de pondération. On trouve par exemple dans la littérature [ASG09], des méthodes permettant de la déterminer par une approche stochastique ou par considération de la fonction produisant les valeurs observées.

2.2. Méthodologie pour la Résolution du problème

2.2.1. Création du nouveau format de données

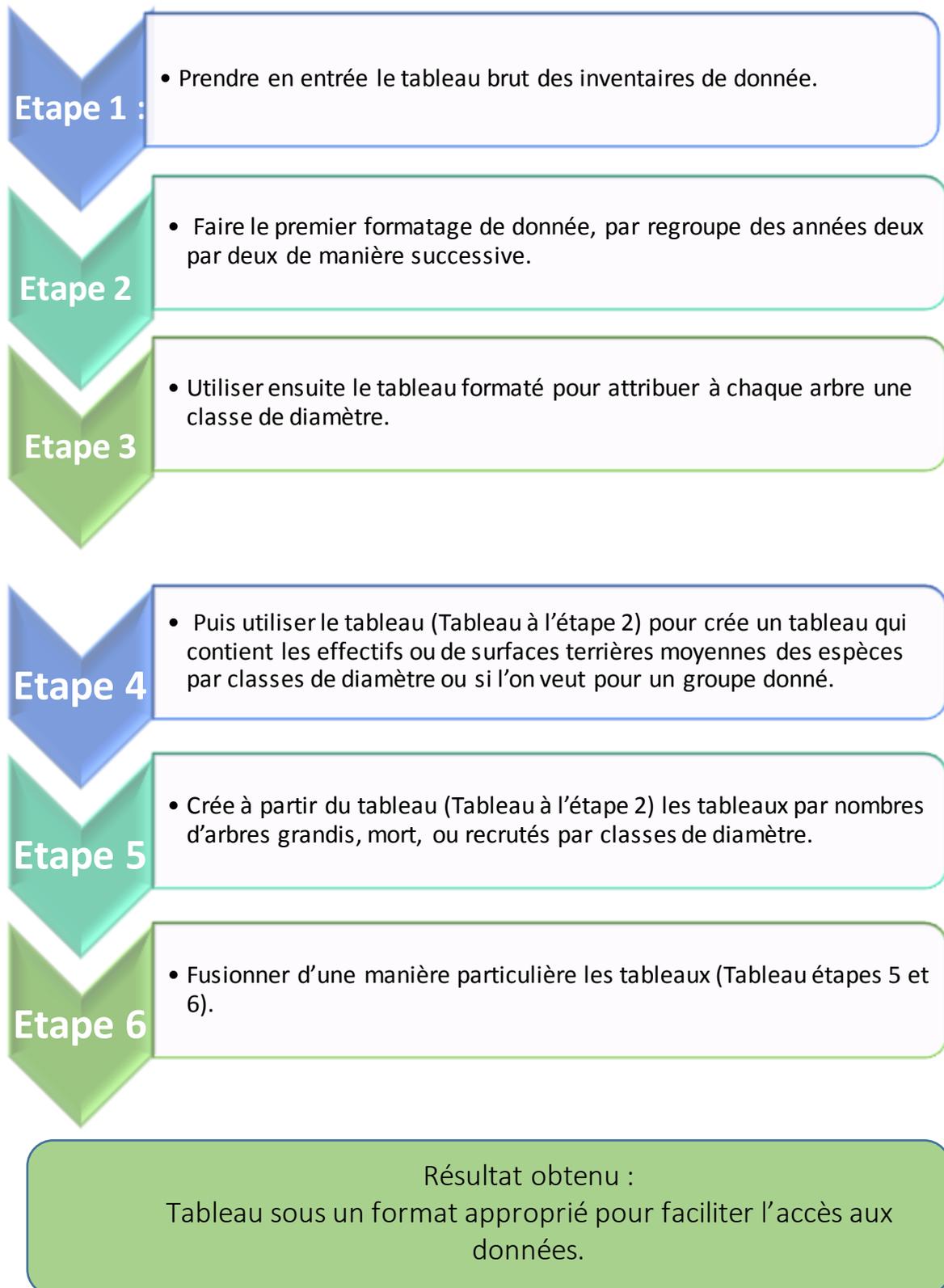
L'inventaire forestier consiste à évaluer la ressource d'une forêt à un moment donné en plus de l'espèce et du diamètre de chacun des arbres inventoriés. Ces informations sont stockées parfois dans un fichier Excel ou autre. Il se pose donc un problème de formatage des données afin de fournir en entrée du cœur du système des données sous le format indiqué.

Pour résoudre ce problème nous avons donc mis sur pied un algorithme propre à notre contexte. Dans un premier temps nous allons présenter l'algorithme proprement dit, puis nous allons faire une représentation de son organigramme afin de permettre une bonne compréhension de l'algorithme.

Proposition d'algorithme : Formatage d'un tableau

But : formater un tableau sous un format adapté à notre contexte.

Données : le tableau d'inventaires forestiers



Organigramme de l'algorithme de formatage de donnée

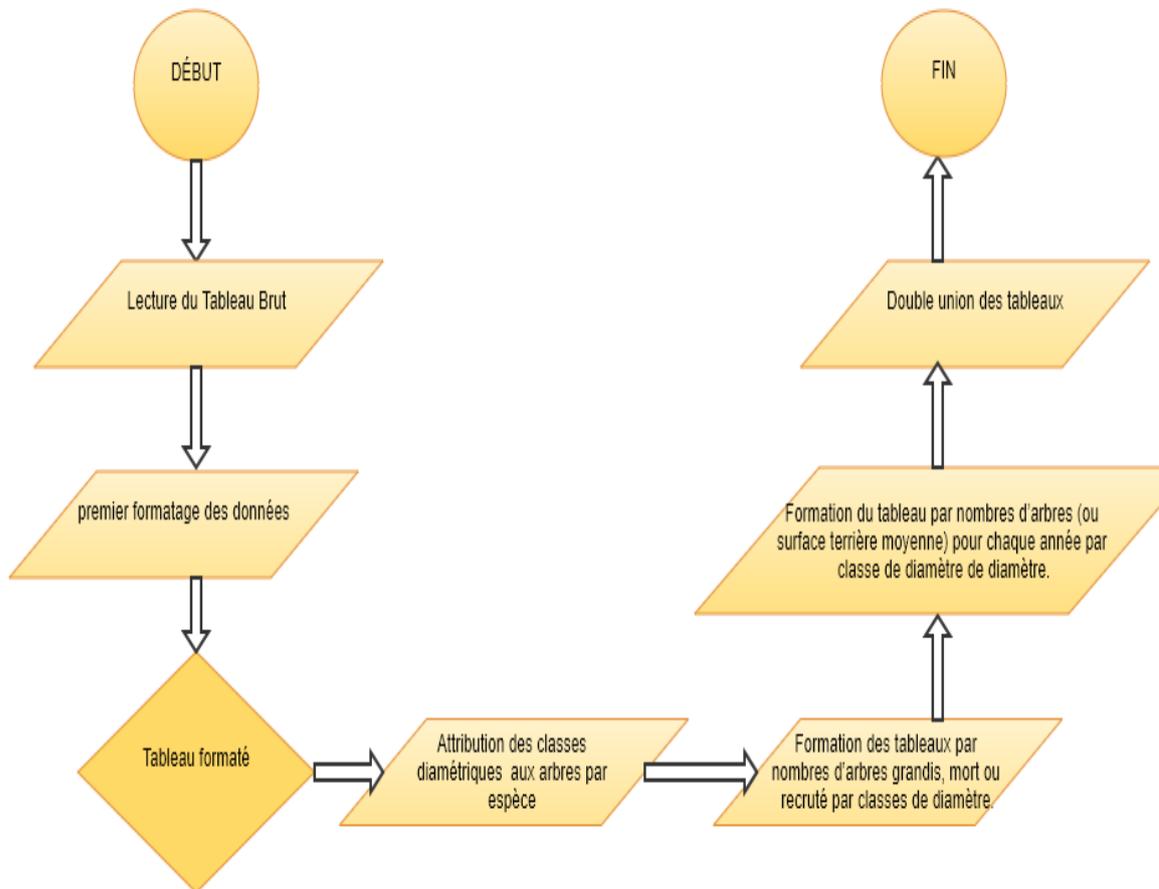


Figure 11 organigrammes de l'algorithme

2.2.3. Méthode de validation des Résultats

Ici il est question pour nous de connaître, et de comprendre les outils que nous utiliserons pour faire des tests et ainsi pouvoir valider nos résultats, et aussi définir la manière avec laquelle nous allons réaliser ces tests.

2.2.3.1. Maximisation de la vraisemblance

Maintenant que nous avons défini la notion de vraisemblance, il semble logique qu'elle varie en fonction du ou des paramètres de la loi théorique en question, spécifiquement dans notre cas la loi de poisson et la loi Binomiale.

Le paramètre à estimer ici est notée : θ « theta » dans la formule (2.18).

En fait, il est plus pratique d'utiliser le **logarithme** de cette fonction (pour remplacer le produit par une somme). Dans la mesure où la fonction $\ln()$ est **croissante**, le **maximum de la fonction de vraisemblance est aussi celui de la log-vraisemblance**.

Une **dérivée partielle** nulle étant un moyen éprouvé pour trouver un extremum, on pose l'équation de vraisemblance. Et on dérive la fonction par rapport au paramètre à estimer [SiteD].

$$\frac{\partial \ln(L(X, \theta))}{\partial \theta} = 0 \quad (2.19)$$

En général, dans les démonstrations, on étudie aussi la **dérivée seconde** pour s'assurer qu'il s'agit bien d'un maximum ; c'est une manière de vérifier si l'estimateur choisi est acceptable [SiteD].

2.2.3.2. Etat de l'art pour les fonctions d'optimisation

Il est important de noter que, en mathématique **l'optimisation** traite de la recherche d'un extremum d'une fonction, dont les entrées peuvent être soumises à des contraintes ;

Nous allons présenter ici trois méthodes d'optimisation numérique utilisées fréquemment sur \mathbb{R} :

Méthode du nombre d'or

Un peu d'histoire

Le nombre d'or est issu d'une relation géométrique permettant d'obtenir un rapport harmonieux dans les arts : peinture, sculpture, musique... . Il a fait son apparition dans la construction des pyramides en Egypte et Grèce où le Parthénon d'Athènes construit au cinquième siècle avant l'an zéro est tel que la relation de ses dimensions est le nombre d'or [CMOP08].

La méthode

La méthode du nombre d'or est une façon simple de trouver **l'unique minimum** d'une fonction à une variable sur un intervalle. $[a, b]$. elle se décrit généralement de la façon suivante :

- Choisir l'intervalle $[a ; b]$ contenant le minimum.
- À chaque itération, on va réduire l'intervalle afin de trouver des intervalles de plus en plus petits contenant toujours le minimum.
- Arrêter les itérations quand l'intervalle : $(b - a)$ est suffisamment petit, c'est à dire quand la longueur de l'intervalle est plus petite que le seuil de tolérance fixé préalablement.

Quand la recherche s'arrête, le milieu du dernier intervalle est une bonne approximation du « vrai » minimum, avec une erreur maximale de $(b - a)/2$ [CMOP08].

Méthode de Newton-Raphson

Un peu d'histoire

La méthode de Newton fut décrite de manière satisfaisante par le mathématicien anglais Isaac Newton (1643-1727) dans De *analysis per aequationes numero terminorum infinitas*. Ce

n'est qu'en 1740 que Thomas Simpson (1710-1761) décrit cette méthode de calcul itératif pour approcher les solutions d'une **équation non linéaire**, utilisant le calcul fluxionnel. Simpson appliqua la méthode de Newton à des problèmes d'optimisation [CMOP08].

La méthode

Si la fonction à minimiser, est deux fois continûment dérivable et que l'on peut les évaluer, nous pouvons donc gérer un algorithme plus rapide que celui de la méthode du nombre d'or. Cette méthode se déroule comme suite [CMOP08].

Soit $f(x)$ la fonction à minimiser :

- Trouver le point qui annule la dérivée premier de $f(x)$ (ie $f'(x_1) = 0$) en commençant par un point initiale bien déterminé x_0 , puis de calculer une meilleure approximation de x_0 .
 - Itérer sur le premier point jusqu'à ce que $f'(x_n)$ est suffisamment proche de 0, ou inférieure à un seuil fixé $|f'(x_n)|$.
- La Méthode de Nelder Mead

Un peu d'histoire

La première méthode de recherche directe a été proposée par Spendley, Hext et Himsworth en 1962. Cet algorithme, appelé méthode du simplexe, n'utilise que deux types de transformations pour former un nouveau simplexe :

- Réflexion
- Rétrécissement

La méthode de Nelder-Mead est encore la méthode la plus populaire, elle est notamment implantée dans plusieurs logiciels (par exemple dans Matlab en 2004 avec la fonction `Fminsearch`) [CMOP08].

La méthode

La méthode originale s'applique à des problèmes sans bornes ; et elle s'arrête quand un minimum local est trouvé.

- Avantages de la méthode
 - L'algorithme de Nelder-Mead est d'ordre zéro, c'est-à-dire, il ne requiert pas de calcul de gradient.
 - Le second avantage de Nelder-Mead est d'être une méthode rapide est robuste, relativement aux méthodes d'ordre zéro.
- Un inconvénient de la méthode classique de Nelder-Mead

Elle peut aboutir à un minimum local. Pour des fonctions ayant plusieurs minima, le point sur lequel l'algorithme converge dépend du point initial, de la taille initiale du simplexe, et des

paramètres de l'algorithme (coefficients d'expansion, de contraction et de réflexion) [CMOP08].

2.2.3.3. Etat de l'art pour des fonctions d'optimisation sur R

La fonction **optimize**

Pour l'optimisation d'une fonction à une variable, la fonction **optimize** () est une variation du nombre d'or.

Syntax : **optimize** (f,intervall, tol=...,...)

La fonction **optimize** () trouve le minimum d'une fonction à seule variable, dans l'intervalle qui le contient.

Elle retourne :

\$minimum : position du minimum si maximum=FALSE

\$maximum : position du maximum si maximum=TRUE

\$objective : la valeur de la fonction à cet endroit [CMOP08].

- Les fonctions d'optimisation en multi dimension

Il existe aussi des fonctions d'optimisation en multi dimension sur R :

La fonction **optim**

Elle permet d'utiliser différentes méthodes d'optimisation, comme Nelder-Mead et Newton-Raphson.

Syntax : **optim**(par, fn, ...)

La fonction **optim**() minimise une fonction objective à plusieurs variables étant donné un point de départ et possiblement une fonction gradient.

Elle retourne :

Par : valeurs initiales pour les paramètres de notre fonction à optimiser

Itérations : nombre d'itérations.

Fn : c'est une fonction de R qui permet d'évaluer la fonction à minimiser. Son premier argument doit être un vecteur de même dimension que **par** ; **optim**() fait appel à cette fonction plusieurs fois afin de trouver le minimum. Elle doit retourner un scalaire [CMOP08].

La fonction **nlm**()

Effectue une minimisation par une méthode de type Newton.

Syntax : **nlm**(fn, par, ...)

Elle retourne :

Estimate : meilleur candidat trouvé

Minimum : valeur de la fonction

Gradient : gradient pour cette valeur

Hessian : hessien pour cette valeur

Code : entier donnant la raison de l'arrêt

Itérations : nombre d'itérations [CMOP08]

Récapitulatif :

Après avoir décrit de manière sommaire les outils que nous envisageons utilisés pour valider nos résultats, il est maintenant question pour nous de décrire la manière avec laquelle nous les ferons.

- Nous allons définir la fonction de vraisemblance pour chaque processus et mettre en place le système de complétion des données.
- Puis nous allons estimer le paramètre θ par la fonction d'optimisation sur R (**optim**)
- Nous le ferons pour un premier temps avec le tableau de donnée au complet (sans irrégularité pour les inventaires de données)
- Ensuite nous le ferons avec le tableau de donnée incomplet (avec des irrégularités pour les inventaires de données).
- Finalement nous le ferons avec le tableau complété par la méthode de complétion.
- Pour terminer nous effectuerons une comparaison entre les valeurs des paramètres obtenus et après analyse nous pourrons tirer des conclusions.

Chapiste 3 : Présentation et Analyse des Résultats

Introduction

L'implémentation de la solution utilise plusieurs technologies, la plupart étant les technologies du langage de calcul scientifique R que nous avons choisi comme langage d'implémentation de la solution compte tenu des contraintes recensés au chapitre 2.

Nous pouvons tout simplement dire que nous disposons d'une base solide pour aborder la phase d'implémentation de cette solution et nous ne saurons la terminer sans bien évidemment présenter le logiciel R et les résultats obtenus.

3.1. Présentation des outils utilisés pour l'implémentation

L'entreprise ayant décidé de faire son propre développement, nous avons opté pour les outils ayant déjà fait leur preuve dans la réalisation d'application statistique, nous avons distingué des logiciels de programmation et de manipulation de données permettant de mettre en œuvre des aides à la décision à savoir R et RStudio.



Figure 12 logo de R et de RStudio

Dans la suite nous allons faire une brève présentation des outils utilisés.

3.1.2. Le logiciel R

Présentation Générale

R est un logiciel Open source (gratuit et libre en développement) développé en C pour permettant de faire des analyses statistiques et de produire des graphiques. Il donne une grande variété pour les calculs statistiques comme la modélisation linéaire ou non linéaire, les tests statistiques, la classification, et autres techniques statistiques et graphiques. Le grand atout de R est sa facilité à produire les graphiques avec des symboles ou des formules mathématiques intégrées. Mais R est également un langage de programmation complet, c'est un clone gratuit du logiciel S-Plus commercialisé par MathSoft et développé par Statistical Sciences autour du langage S.



Figure 14 logo logiciel R

3.2.2. Le logiciel RStudio

Présentation Générale

RStudio est un environnement de développement intégré (IDE) pour R. Il comprend une console, un éditeur de mise en surbrillance de syntaxe qui prend en charge l'exécution directe de code, ainsi que des outils de traçage, d'historique, de débogage et de gestion d'espace de travail.

RStudio est disponible dans les éditions open source et commerciales et s'exécute sur le bureau (Windows, Mac et Linux) ou dans un navigateur connecté à **RStudio** Server ou **RStudio** Server.



Figure 13 Logo de RStudio

Interface graphique de RStudio

En informatique, une interface graphique (en anglais GUI pour graphical user interface) est un dispositif de dialogue homme-machine, dans lequel les objets à manipuler sont dessinés sous forme de pictogrammes à l'écran, de sorte que l'utilisateur peut utiliser en imitant la manipulation physique de ces objets avec un dispositif de pointage, le plus souvent une souris.

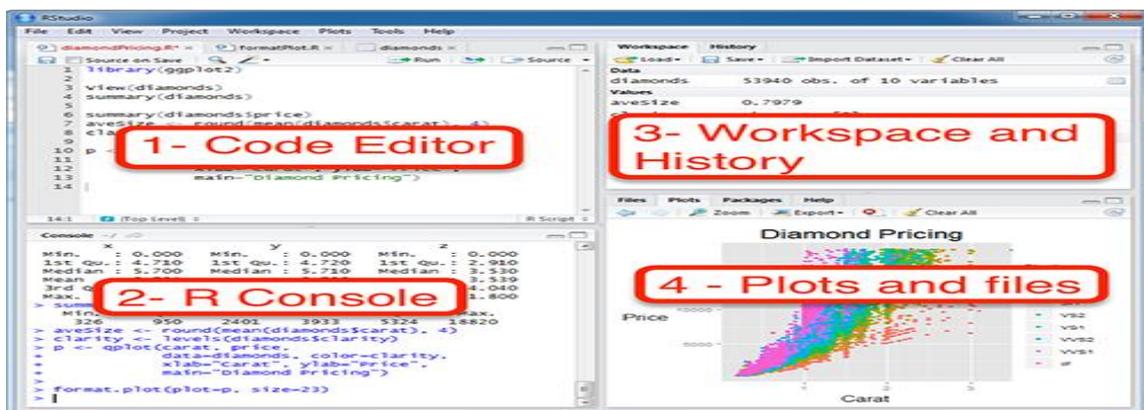


Figure 16 Editeur RStudio

Récapitulatif :

Nous venons de présenter les deux logiciels suffisants et nécessaires pour notre travail, à savoir **RStudio** et le langage R, **RStudio** nous servira d'éditeur pour les scripts (programme sur fichier) et le langage R qui nous permettra de faire l'application des statistiques. La réunion de ces logiciels nous conduira donc à l'élaboration d'un modèle statistique.

3.3. Présentation des Résultats obtenus

Dans cette partie, nous présenterons

- L'échantillon de test
- Les valeurs des paramètres estimés dans les différents cas de figures.
- La précision des estimateurs.
- Information disponible

3.3.1. Echantillon de test

3.3.1.1 Critères de choix de l'échantillon

Pour effectuer les tests nous nous sommes fixés un échantillon limité, en effet vu la quantité de donnée et l'ensemble des instructions qui constituent respectivement les données en entrée et le corps de notre algorithme, nous avons réduit la taille de l'échantillon pour les raisons suivantes :

- Faible capacités de calculs.
- Grande quantité de données.

Nous présentons par la suite les critères sur lesquelles nous avons effectué la réduction de la taille des données :

- La classe d'arbre qui contient le plus d'arbre.
- Les espèces d'arbre qui sont le plus prisés par les exploitants forestiers
- Les parcelles qui sont les plus représentatifs.

3.3.1.2 Présentation de l'échantillon

On représente les données de la manière suivante :

Pour la croissance

- Première partie du tableau

Cette partie du tableau fournit des informations sur le nombre d'arbres d'une classe de diamètre précise qui ont grandi pendant une période donnée par rapport à un nombre d'arbres observé l'année ou l'inventaire a été effectué.

Tableau 4 première partie du tableau de donnée pour la classe [9-20]

Zone	Année	Espèce	Période	Nombre d'arbre grandis	Nombre d'arbre observés
111	1986	1	1	1	4
112	1986	2	1	0	3
113	1988	3	2	0	2
114	1989	4	2	1	9
121	1994	5	1	0	8
122	1991	6	3	0	1
124	1989	7	1	0	6
131	1993	8	2	0	4
133	1992	9	4	2	11
...
244	1995	153	2	1	5

○ Deuxième partie du tableau

Cette deuxième partie du tableau fournit des informations sur le nombre d'arbres (ou la surface terrière moyenne) des arbres d'une classe de diamètre précise, l'année où l'inventaire a été effectué.

Tableau 5 deuxième partie du tableau pour la classe [9-20]

Zone	X					
111	1986	x_0^{1986}	x_1^{1986}	x_2^{1986}	...	x_8^{1986}
112	1987	x_0^{1987}	x_1^{1987}	x_2^{1987}	...	x_8^{1987}
113	1988	x_0^{1988}	x_1^{1988}	x_2^{1988}	...	x_8^{1988}
114	1989	x_0^{1989}	x_1^{1989}	x_2^{1989}	...	x_8^{1989}
121	1990	x_0^{1990}	x_1^{1990}	x_2^{1990}	...	x_8^{1990}
122	1991	x_0^{1991}	x_1^{1991}	x_2^{1991}	...	x_8^{1991}
124	1992	x_0^{1992}	x_1^{1992}	x_2^{1992}	...	x_8^{1992}
131	1993	x_0^{1993}	x_1^{1993}	x_2^{1993}	...	x_8^{1993}
133	1994	x_0^{1994}	x_1^{1994}	x_2^{1994}	...	x_8^{1994}
244	1995	x_0^{1995}	x_1^{1995}	x_2^{1995}	...	x_8^{1995}
258	1996	x_0^{1996}	x_1^{1996}	x_2^{1996}	...	x_8^{1996}
...
220	1986	x_0^{1986}	x_1^{1986}	x_2^{1986}	...	x_8^{1986}

Récapitulatif sur l'échantillon :

Nous avons donc décidé de limiter l'échantillon sur une période de 10 ans de 1986 à 1996 car pendant cette période les données ont été récoltées de manière régulière ; sur exactement 09 espèces car ces espèces sont majoritairement représentées dans l'ensemble de données et à 09 zones ou parcelles de terrain dans lesquelles nous trouvons ces arbres.

3.3.2 Valeurs des paramètres estimés

Nous présenterons les résultats qui sont ici la détermination des paramètres inconnus $\theta = (\theta_0 \theta_1 \dots \theta_n)$ avec $n \in \{1, 2, 3 \dots, 10\}$ en plusieurs phases à savoir :

Résultats obtenus avec :

- Les données au complet θ :

Tableau 6 Données du paramètre θ initial

θ	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
valeur	-2.301	0.081	0.016	-0.063	-0.173	0.160	1.365	1.421	3.723	0.805

- une année d'absence d'inventaire

Lorsque l'année n'est pas complétée nous obtenons θ_{-1} :

Tableau 7 paramètre θ pour moins un an

θ_{-1}	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Valeur	0.487	0.111	-0.154	-0.061	0.373	1.402	1.072	1.253	1.928	-0.026

Puis après avoir complété les données par la méthode d'interpolation de l'année manquante nous obtenons θ_{+1} :

Tableau 8 paramètre θ pour l'année complétée

θ_{+1}	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
valeur	-1.139	0.13 0	-0.039	-0.232	0.343	0.727	-0.388	2.193	1.352	0.891

- deux années d'absence d'inventaire

Lorsque les deux années ne sont pas complétées nous obtenons θ_{-2} :

Tableau 9 paramètre θ pour moins deux ans

θ_{-2}	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Valeur	4.702	0.091	-0.048	-0.307	0.345	1.952	-1.362	0.048	1.414	-2.089

Après avoir complété les données par la méthode d'interpolation des années manquante nous obtenons θ_{+2} :

Tableau 10 paramètre θ pour deux années complétés

θ_{+i}	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Valeur	0.616	0.188	0.056	-0.487	0.227	1.222	-1.367	0.715	1.409	1.182

- Trois années d'absence d'inventaire

Lorsque les trois années ne sont pas complétées nous obtenons θ_{-3} :

Tableau 11 paramètre θ pour deux ans complétés

θ_{-3}	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Valeur	1.683	-0.026	-0.192	0.240	0.317	0.656	1.228	-0.448	1.062	0.711

Après avoir complété les données par la méthode d'interpolation des années manquante nous obtenons θ_{+3} :

Tableau 12 paramètre θ pour trois années complétés

θ_{+3}	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Valeur	0.187	0.030	-0.167	-0.048	0.459	2.013	-0.193	0.865	1.577	0.747

3.3.2.1 Précision des estimations

Pour étudier la précision de nos estimateurs, nous allons considérer dans un premier temps l'erreur absolue et l'erreur quadratique moyenne (RMSE) Cette erreur nous donne une mesure synthétique de l'erreur globale dans une seule valeur.

3.3.2.1.1 L'erreur absolue

Il s'agit des erreurs individuelles, en effet elle est le résultat de la valeur absolue de la différence entre la valeur estimée de la grandeur que l'on estime et de la valeur référence que nous avons choisie comme une bonne.

Il faut noter ici que ; L'**erreur absolue** mesure l'imprécision sur une mesure que nous effectuons.

Tableau 13 Erreur absolue

	Vréel	Pour un an		Pour deux ans		Pour trois ans	
	θ	E_1	E_2	E_3	E_4	E_5	E_6
Erreur absolue	θ	$ \theta - \theta_{-1} $	$ \theta - \theta_{+1} $	$ \theta - \theta_{-2} $	$ \theta - \theta_{+2} $	$ \theta - \theta_{-3} $	$ \theta - \theta_{+3} $
θ_0	-2.301	2.789	1.162	7.003	2.918	3.984	2.489
θ_1	0.081	0.030	0.049	0.010	0.106	0.107	0.050
θ_2	-0.063	0.091	0.024	0.014	0.119	0.129	0.104
θ_3	-0.173	0.111	0.059	0.134	0.314	0.413	0.124
θ_4	0.160	0.212	0.183	0.185	0.066	0.157	0.298
θ_5	0.762	0.639	0.034	1.190	0.459	0.105	1.251
θ_6	1.365	0.293	1.754	2.728	2.732	0.137	1.559
θ_7	1.421	0.168	0.771	1.373	0.705	1.870	0.556
θ_8	3.723	1.794	2.370	2.308	2.313	2.660	2.145
θ_9	0.805	0.832	0.086	2.895	0.376	0.094	0.058

Analyse des résultats

Pour faciliter l'interprétation des résultats nous avons réalisé les graphes suivants :

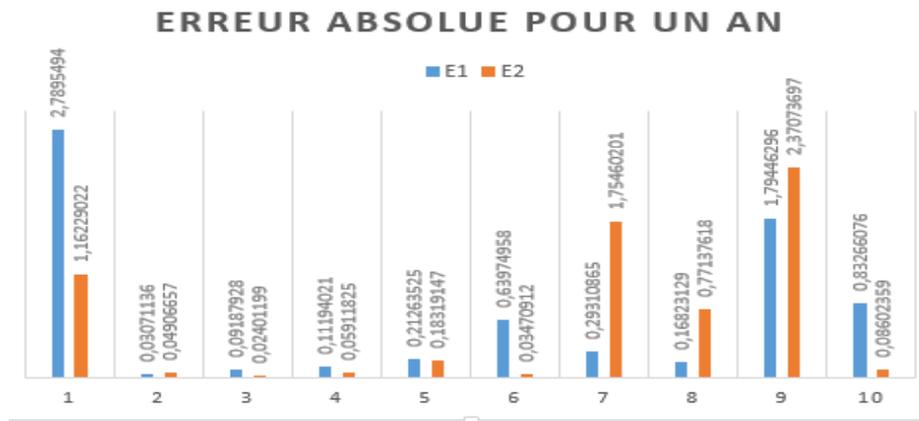


Figure 14 Erreur absolue pour un an

L'erreur individuelle dans le cas des données complétées pour les 5 premières valeurs et pour la dernière valeur est nettement inférieure comparée au cas où les données sont absentes.

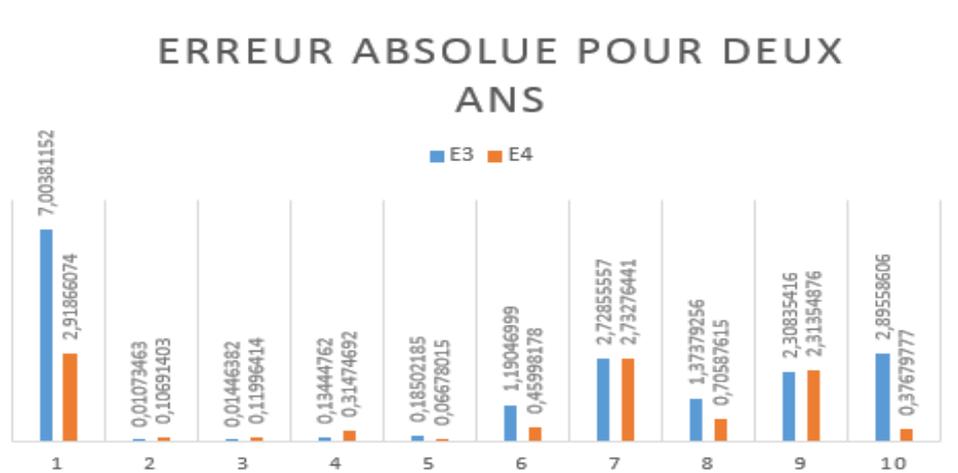


Figure 15 Erreur absolue pour deux ans

Nous constatons ici que l'erreur individuelle dans le cas des données complétées pour la 1^{er}, 3^{ème}, 4^{ème}, 5^{ème}, 6^{ème}, la 8^{ème} et la dernière valeur est nettement inférieure comparé au cas des données absentes.

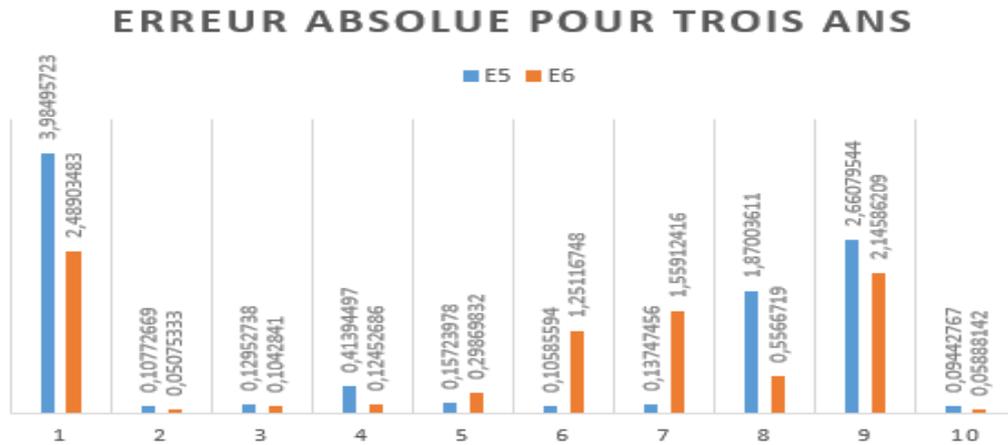


Figure 16 Erreur absolue pour trois ans

Constat fait, ici l'erreur individuelle dans le cas des données complétées pour les quatre premières données, des 8^{ième}, 9^{ième} et de la dernière valeur est nettement inférieure par rapport au cas des données absentes.

Récapitulatif

D'après les résultats obtenus nous pouvons en effet affirmer que la méthode qui consiste à compléter les données manquantes, réduit de manière significative l'imprécision des estimations obtenues pour les valeurs inconnues.

3.3.2.1.2 Erreur quadratique moyenne

En considérant les paramètres obtenus lorsque les données étaient complètes comme les vraies valeurs des paramètres à estimées soit :

Pour calculer l'erreur quadratique moyenne RMSE⁵, les erreurs individuelles sont tout d'abord élevées au carré, puis additionnées les unes aux autres. On divise ensuite le résultat obtenu par le nombre total d'erreurs individuelles, puis on en prend la racine carrée.

Tableau 14 formule de la RMSE

$$\text{RMSD}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})} = \sqrt{\text{E}((\hat{\theta} - \theta)^2)}.$$

⁵ root mean square error

Tableau 15 Tableau des RMSE

RMSE			θ	
1	S1	E1	θ_{-1}	1.108406
2		E2	θ_{+1}	1.034099
3	S2	E1	θ_{-2}	2.712351
4		E2	θ_{+2}	1.493997
5	S3	E1	θ_{-3}	1.634632
6		E2	θ_{+3}	1.234011

Il est important de noter que La propriété la plus désirable pour un estimateur est d'avoir une faible Erreur Quadratique Moyenne (ce qui n'exige pas forcément d'être sans biais⁶) soit la figure suivante :

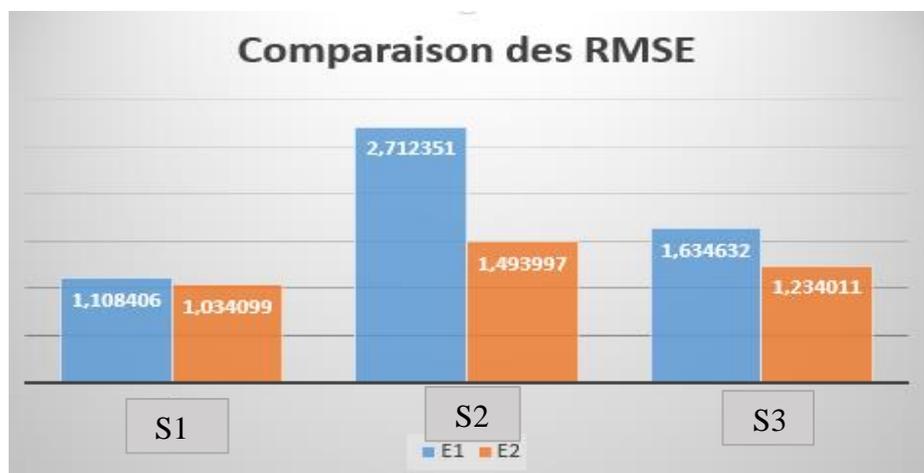


Figure 17 Comparaison des RMSE

Les résultats obtenus nous permettent de conclure que, la méthode d'interpolation des données manquantes influence considérablement la précision des paramètres estimés.

⁶ Le **biais** d'un estimateur est la différence entre la valeur de son espérance et la valeur de la variable aléatoire qu'il est censée estimer.

3.3.3 Récapitulatif

Par rapport au test effectué précédemment nous pouvons déduire le tableau suivant :

Tableau 16 Récapitulatif

Tableau comparatif	Algorithme avec complétion	Algorithme sans complétion
Qualité d'approximation de θ	Très bonne	faible
complexité en temps	Grand	moyen
Complexité en mémoire	moyen	petit

En somme, on peut dire que notre algorithme avec complétion de donnée et celui sans complétion de donnée, font la différence au niveau de la précision des estimations et de la complexité en temps, nous pouvons expliquer d'une certaine façon pourquoi en terme de temps l'algorithme sans complétion de donnée est légèrement supérieure ; en effet pour notre algorithme, nous effectuons certains calculs en plus pour compléter les données absentes. L'utilisation des inventaires de données aux complets, nous rassure de notre développement et permet de valider notre démarche par la concordance que nous avons dans les résultats. Ces résultats obtenus nous permettent d'étendre le champ d'application du logiciel DafSim au cas d'inventaires à intervalles de temps irréguliers. Pour ce qui est de notre algorithme, on comprend que la recherche de la bonne solution augmente le temps d'exécution.

Conclusion

Nous sommes parvenus au terme de notre étude. Il était question de mettre en place un module de calibration des paramètres du modèle **DafMod**, basé sur le suivi des processus de dynamique forestière. L'analyse des modèles des processus de dynamique forestière nous a fait ressortir deux modèles de régression : le modèle de régression logistique permettant d'inférer sur la probabilité que les arbres ont de grandir ou de mourir en fonction de leur diamètre, et la régression de Poisson qui permet d'inférer sur la régénération en ne considérant que le recrutement d'un arbre. Après avoir présenté le modèle du simulateur, les hypothèses de simulation, le modèle de classe de diamètre et la notion de vraisemblance ; nous avons proposé des modélisations et défini les processus de : croissance, recrutement, et de mortalité des arbres à la fin desquels nous avons proposé des modèles se rapprochant le plus possible aux processus directeurs de la dynamique forestière. Par la suite nous avons présenté la méthode d'interpolation des données de variables explicatives qui nous a fait aboutir à la présentation d'une méthodologie pour la résolution du problème. Ce dernier cas permettant d'arriver à notre objectif de généralisation à des inventaires espacés d'une manière quelconque. Dans le cas du processus de croissance, l'implémentation de notre modèle pour les inventaires en intervalle de temps irrégulier nous a permis de mettre en place un module fiable et prometteur. En guise de perspective, il serait intéressant d'être capable de collecter les données sur les arbres (leurs surfaces terrières) à l'aide d'un satellite qui estimera par des images les informations nécessaires au fonctionnement du logiciel DafSim ce qui réduira certainement la probabilité qu'un inventaire prévu ne soit pas réalisé. Il s'agira de louer un Satellite, de l'orienter vers la zone désirée, d'effectuer des captures d'images et enfin analyser ses captures pour en tirer des informations.

BIBLIOGRAPHIE

[**CMUC17**] : Christophe Chesneau ; Modèle de Régression, Université de Caen 2017

[**TEM2017**]: TEGUIA, Mémoire de fin d'étude pour l'obtention du diplôme d'ingénieur de conception, 2017

[**ASG09**] : Avram Florin. Statistique spatiale, géostatistique, régression et interpolation. Decembre 2009.

[**VRFM2010**] : Vivien ROSSI Fabien Campillo, Frederic Mortier. Chaines de markov modèles matriciels. Mars 2010

[**SiteC**]:<http://statwww.epfl.ch/davison/teaching/ProbStatSC/20032004/Chapitre8.pdf>

[**SiteD**] : <http://www.jybaudot.fr/Inferentielle/vraisemblance.html>

[**UMB07**] : Usher M.B A matrix model for forest management. Université Montpellier II. Juin 1969.

[**DKPD11**] : Dakis-Yaoba ouedraogo. Prédiction de la dynamique forestière à l'aide d'un modèle matriciel qui incorpore la variabilité de la réponse des espèces à l'environnement, PhD thesis, Université Montpellier II, 2011.

[**CMOP08**]: Carbone Mathieu, Poma Elodie ; Stage M1 optimisation numérique sur le logiciel R, 2007-2008.

[**MHF09**] Zhani Mohamed, Faten, Elbiaze Halima, and Kamoun Farouk. Analysis and prediction of real network traffic. Novembre 2009. Conception d'un réseau local.

[**Moh15**] ABIED Mohamed, Ismail. La technologie wdm. Octobre 2015. SUPINFO International University.

[**BUT**] Besse Philippe and Laurent Beatrice. Apprentissage Statistique : modélisation, prévision et datamining. Université de Toulouse/Département de Génie Mathématique et Modélisation.

[**GPCSUP**] Gambette Philippe. Classification supervisée et non supervisée. Université Paris-Est Marne-la-Vallée/Ingénierie Linguistique, 2014.

[**HCS**] Hadd Mustapha. Classification de la population en catégories socio-économiques : méthodologie et application pratique, 1999. www.memoireonline.com.

[**AUP15**] AMAJDA. Rapport de recherche sur le réseau de kohonen, fev 2015. Université Paris 1-Sorbonne.

[**BCUL09**] Biemacki Christophe. Pourquoi les modèles de mélange pour la classification ? CNRS et Université de Lille 1, Villeneuve d'Ascq, France, 2009.

[**DL77**] Dempster, Laird, and Rubin. Maximum likelihood from incomplete data via the EM algorithm,

[**SFP09**] Santos Frédéric. L'algorithme em : une courte présentation. CNRS, UMR 5199 PACEA, août 2009.

[**LMTA10**] Lejeune Michel. Statistique La théorie et ses applications. Springer, 2010. Deuxième édition.