Aspects statistiques de la stabilité en dynamique des populations : application au modèle de Usher en foresterie.

Mélanie Zetlaoui

N° d'ordre : 8503

UNIVERSITÉ PARIS XI

THÈSE

présentée pour obtenir

LE TITRE DE DOCTEUR EN SCIENCES

Spécialité Mathématique

par
Mélanie Zetlaoui

Aspects statistiques de la stabilité en dynamique des populations : application au modèle de Usher en foresterie.

Soutenue le 7 décembre 2006 devant la commission d'examen :

M.	Avner Bar-Hen	Directeur de thèse
M.	Patrice Bertail	Rapporteur
M	Alain Franc	Examinateur
Mme.	Elisabeth Gassiat	Présidente du jury
M.	Nicolas Picard	Co-directeur de thèse
M.	Jérôme Saracco	Rapporteur

Remerciements

Mes premiers remerciements vont tout droit à mes directeurs de thèse. Avner Bar-Hen m'a dirigée et soutenue tout au long de ma thèse. Ses conseils m'ont été d'une aide précieuse dans l'aboutissement de mon travail. Nicolas Picard, par les discussions que j'ai eues avec lui et par ses remarques toujours très précises, m'a permis d'approfondir ma réflexion sur mon travail de recherche. Sa grande disponibilité et son attention ajoute à ses qualités de chercheur.

Je tiens à remercier également tous les membres du Jury. Merci à Patrice Bertail et Jérôme Saracco, pour avoir accepté d'éplucher ce mémoire et pour l'intérêt qu'ils ont porté à mon travail. Leurs remarques et les discussions qu'elles ont amenées m'ont été particulièrement enrichissantes. Merci aussi, pour avoir accepté d'évaluer ce travail, à Alain Franc et à Elisabeth Gassiat, que je remercie tout particulièrement pour avoir suivi mes travaux tout au long de ma thèse et pour son soutien.

Grand merci aux chercheurs du laboratoire OMIP de l'INAPG pour m'avoir accueillie, et tout particulièrement à Jean-Jacques Daudin ainsi qu'à Etienne Klein pour avoir suivi mon travail lors de mes comités de thèse et dont les remarques pertinentes m'ont été précieuses.

Ce travail de thèse s'est faite en collaboration avec le CIRAD-Forêt, d'où est parti mon sujet de thèse. Merci en particulier à Sylvie Gourlet-Fleury pour m'avoir fait partager son expérience sur les questions forestières, et à ceux qui ont rendus mes séjours dans ce laboratoire agréables, en particulier Frédéric, Jean-Marc et Phillippe.

Durant ma thèse, j'ai été rattachée, en tant qu'ATER, au laboratoire de Mathémathiques de Paris XIII dans un premier temps. Je tiens à remercier tout particulièrement Francesco Russo pour ses encouragements dans mon travail de recherche. J'ai ensuite poursuivi mon service au laboratoire de Mathématiques d'Evry-Val d'Essonne. Je remercie Monique Jeanblanc et Bernard Prum pour leur accueil, et tous ceux avec qui j'ai eu le plaisir d'enseigner.

Dans la dernière année de ma thèse, j'ai côtoyé les membres de l'unité

Mét@risk de l'INRA. Merci à eux pour les bons moments passés ensemble, et particulièrement à Stéphan Clémençon pour s'être intéressé à mon travail et m'avoir fait partagé son expérience de la recherche et ses connaissances.

Merci aussi aux thésards d'Orsay du bâtiment 430 et particulièrement à mes amis, Cristian, Ana, Sophie, Nati, Hector, et Jean-Maxime.

Je remercie enfin tous mes amis, et particulièrement Boris pour son écoute et ses conseils rassurants, Clem pour sa bonne humeur, ma Coco que je retrouve toujours, malgré de longues séparations, comme si c'était hier, et Fred pour son affection.

Ma pensée va enfin à mes parents pour leur soutien, leurs encouragements et leur confiance de toujours, et à ma Nono pour ce qu'elle est tout simplement.

Résumé

Les modèles matriciels sont souvent utilisés pour prédire l'évolution en temps discret d'une population structurée par âge ou par taille. Le modèle de Usher est un modèle matriciel qui décrit l'évolution des individus suivant leur taille et restreint les transitions entre les classes d'état. Il est particulièrement adapté pour décrire la dynamique d'un peuplement forestier et sert de guide dans la gestion des forêts. Cette étude porte sur les prédictions dans l'état stationnaire du modèle. L'objectif principal est la construction d'intervalles de confiance de ces prédictions.

Le premier chapitre fait un état de l'art sur le modèle de Usher et pose la problématique de la thèse.

Le deuxième chapitre est consacré à la construction d'intervalles de confiance des prédictions. Pour cela la distribution asymptotique des estimateurs du maximum de vraisemblance des prédictions est obtenue grâce à la delta-méthode. Des simulations permettent de vérifier la validité des résultats asymptotiques pour différentes tailles d'échantillon.

Dans le troisième chapitre, les intervalles de confiance asymptotiques sont affinés en cherchant des estimateurs robustes des paramètres du modèle. Cette recherche est guidée par deux types de contraintes du modèle portant sur sa structure discrète et sur la dynamique de la population. Les estimateurs des paramètres ainsi construits sont des L-estimateurs exprimés dans un modèle statistique multidimensionnel. Le critère de robustesse utilisé est la sensibilité des estimateurs, basé sur la notion de fonction d'influence. L'utilisation de la fonction d'influence permet de plus de déterminer le comportement asymptotique des estimateurs et d'en déduire des intervalles de confiance.

Le quatrième chapitre étend les résultats du deuxième chapitre au cas du modèle densité-dépendant, dans lequel les paramètres sont fonctions des caractéristiques courantes de la population. L'existence et l'unicité du vecteur de distribution stationnaire sont au préalable vérifiées.

Les résultats théoriques sont appliqués un jeu de données réelles d'un peuplement forestier en Guyane Française et les implications pratiques sont discutées.

MOTS-CLÉS : modèle de population, modèle matriciel, statistique asymptotique, robustesse, fonction d'influence, sensibilité, L-estimateurs, dynamique forestière.

Abstract

Matrix models are often used to describe the discrete-time evolution of a age-structured or size-structured population. The Usher model is a matrix model describing a size-structured population that is characterised by a restriction on the transitions between the state classes. It is well adapted to describe the dynamic of a forest stand and is used to deal with forest management. This study turns on predictions in the stationary state of the model. The main object is the construction of confidence intervals of these predictions.

The first chapter gives an overview on the Usher model and lines out the problematic of the thesis.

The second chapter addresses the construction of asymptotic confidence intervals of predictions. Therefore, the asymptotic distribution of the maximum likelihood estimator of predictions is obtained by the delta method. Simulations allow to verify the validity of asymptotic results for different sample sizes.

In the third chapter, the confidence intervals are refined by searching robust estimators of model parameters. The construction of these estimators respects the model constraints concerning its discrete structure and the dynamic of the population. The parameters estimates are L-estimators expressed in a multidimensional statistical model. The robustness criteria used is the estimator's sensibility based on the influence function. The asymptotic behaviour of the estimators is moreover determined by using the influence function and hence confidence intervals are derived.

A fourth chapter extends the results of the second chapter to the more general density-dependant Usher model, where the parameters depend on the varying characteristics of the population during time.

The theoretical results are applied on a real data set of a forest stand in French Guyana and the practical implications are discuss.

KEY WORDS : population model, matrix model, asymptotic statistics, robustness, influence function, sensibility, L-estimators, forest dynamic.

AMS-Classification : 62F10, 62E20, 62F35, 62P12.

Table des matières

Introduction

Un modèle mathématique est une représentation d'une situation réelle donnée à l'aide d'objets mathématiques. Il nécessite de faire des hypothèses au préalable afin d'avoir un modèle traduisible de façon simple en termes mathématiques, et d'utiliser des outils mathématiques permettant de décrire la réalité. Il permet ainsi de rendre intelligible un phénomène réel en en donnant une description la plus fidèle possible.

Il existe principalement deux types de stratégie dans la construction d'un modèle (Legay, 1997 ; Pavé, 1994) :

- on peut premièrement s'inspirer de lois générales (par exemple des lois de la physique) faisant appel à peu d'hypothèses. Dans ce cas, la pertinence du modèle doit être vérifiée en la confrontant à des situations expérimentales, ce qui permet d'éviter des conclusions hasardeuses ;
- la deuxième approche est de partir d'une situation concrète dont on a une description plus ou moins complète. Le modèle doit alors se baser sur des hypothèses contraignantes, imposées par le phénomène réel étudié. La construction du modèle est alors plus complexe que dans l'approche précédente, mais le modèle obtenu a plus facilement un sens par rapport à la situation réelle étudiée.

En pratique, les deux stratégies ne s'opposent pas forcément, et peuvent être utilisées simultanément dans la modélisation d'une situation.

L'élaboration d'un modèle nécessite au préalable de définir très clairement la situation concrète et ensuite de choisir les hypothèses :

- pour définir la situation concrète, on doit effectuer une analyse de cette situation comprenant les objets ou le phénomène à représenter ainsi que le cadre dans lequel ils sont étudiés, les données expérimentales et les connaissances disponibles. Elle permet en particulier de définir les variables et guide dans le choix des relations entre ces variables. Ceci signifie de plus que le problème soit posé le plus précisément possible et suivant un certain point de vue (écologique, médical, économique...) ;
- le choix des hypothèses est guidé par les informations expérimentales disponibles et les caractéristiques des objets étudiés d'une part, et par la

simplification du modèle mathématique et des développements mathématiques qu'ils impliquent d'autre part. Ce choix est déterminant dans les résultats obtenus : des hypothèses très simplificatrices par exemple peuvent conduire à un résultat peu réaliste. L'analyse des résultats se fera alors en relation avec les hypothèses posées.

Ce travail de thèse s'inscrit dans la seconde stratégie de la modélisation, dans la mesure où les équations mathématiques développées émanent de questions biologiques relatives au fonctionnement de l'écosystème forestier. Au travail de manipulation des objets mathématiques s'ajoute donc, en amont, un travail pour formaliser les questions biologiques en termes mathématiques adéquats. Le choix des objets mathématiques étudiés n'est pas fait en fonction de leur intérêt mathématique, mais en fonction de leur adéquation à la question biologique posée. Ce choix n'est pas forcément immédiat ni facile : il nécessite de comprendre la question biologique posée, donc de s'approprier le contexte biologique du problème, et il nécessite de maîtriser les différentes alternatives possibles pour répondre à la question.

Précisons à présent le contexte biologique du problème qui a motivé cette thèse.

La nécessité des modèles pour la gestion des forêts

Le contexte biologique de cette thèse est la gestion des forêts naturelles tropicales humides (Bergonzini and Lanly, 2000). Le mode de gestion de ces forêts, à la différence des forêts tempérées, minimise l'intervention humaine : les forêts sont exploitées, puis laissées à elles-mêmes pendant un temps suffisamment long (appelé la rotation) pour que la ressource se renouvelle naturellement. Une gestion est dite durable si le cycle exploitation–repos peut être maintenu sur le long terme sans qu'il y ait dégradation du stock de bois exploitable (Estève, 2001 ; Guéneau, 2006). La clé d'une gestion durable est l'équilibre entre les prélèvements et le renouvellement naturel de la ressource à l'issue d'une rotation. La recherche de cet équilibre peut être formalisée comme un problème à trois variables : le diamètre minimum d'exploitation de chaque espèce, la durée de rotation et le stock de bois reconstitué à la fin de chaque rotation (Sist et al., 2003). La résolution de ce problème nécessite de prédire la dynamique de la forêt. Cette prédiction requiert l'utilisation de modèles de dynamique de population (Vanclay, 1994).

Différents types de modèles en dynamique des populations

Un peuplement forestier régulier, c'est-à-dire monospécifique équienne, est souvent résumé par la densité d'arbres et par son arbre moyen, sup-

posé représentatif de tous les arbres du peuplement. Cet arbre moyen est décrit par un certain nombre de variables comme la hauteur, la circonférence de l'arbre et l'âge. Des lois biologiques empiriques permettent d'avoir des modèles simples mettant en relation ces variables. En forêt hétérogène (c'est-à-dire plurispécifique ou inéquienne), ces lois ne sont plus valables. Pour prendre en compte l'hétérogénéité, on considère des modèles plus sophistiqués tels que les systèmes dynamiques, qui vont décrire la trajectoire du peuplement au cours du temps. Ces modèles reposent sur différents niveaux de description : l'individu, le groupe d'individu et la population. Ces différents niveaux de description conditionnent le type de modèle que l'on développera.

Un premier type de modèle repose sur la description de la population par une fonction de distribution sur une ou plusieurs variables. Ces modèles ont plusieurs origines :

- en écologie, par Leslie en 1945 en démographie humaine où la variable est l'âge ;
- en génétique des populations, au début du XX^e siècle, suite aux travaux de Hardy, Fisher, Haldane et Wright, dans le cadre de la modélisation de l'évolution des fréquences alléliques dans une population ;
- ils sont également liés aux modèles issus de la physique statistique, dont l'origine remonte à Maxwell et Boltzman au XIX^e siècle.

Dans ces familles de modèles, une population est résumée par une fonction de distribution sur une ou plusieurs variables. Cette modélisation consiste à suivre l'évolution dans le temps de la fonction de distribution. Le modèle stochastique sous-jacent décrit les trajectoires individuelles par une chaîne de Markov lorsque le temps est discret et, plus généralement, par un processus stochastique lorsque le temps est continu. En foresterie, la fonction de distribution peut être la distribution diamétrique, la distribution des arbres selon leur hauteur, etc. Dans cette classe de modèles, on différencie les modèles selon que la fonction de distribution et le temps sont des variables discrètes ou continues. Les modèles les plus simples sont les modèles matriciels qui représentent l'évolution des distributions en états et en temps discrets. On s'intéresse dans cette thèse à ces modèles. Ce choix sera justifié dans le prochain chapitre. Ils peuvent être également étendus au cas où l'état et/ou le temps sont continus. Si les deux sont continus, le modèle est alors défini par des équations aux dérivées partielles. Ces modèles ont été principalement introduits par McKendrick en 1926 et redécouvert en 1959 par von Foerster en démographie humaine, le modèle de McKendrick-von Foerster reposant sur l'équation de renouvellement de Lotka (1939) (Franc et al., 2000 ; Caswell, 2001).

Un deuxième type de modèle est le modèle individu-centré (Huston et al.,

1988 ; DeAngelis and Gross, 1992), où la description du peuplement se fait au niveau des individus, dépendant ou non des distances. Les modèles individuels indépendants des distances ont été introduits en foresterie dès les années 1950 (Staebler, 1951 ; Newnham, 1964). Ils sont en fait redondants avec les modèles de distribution, puisqu'en grande population, ils tendent vers des modèles de distribution continue. Mais, historiquement, ce sont les premiers modèles individuels à être étudiés. Par la suite se sont développés les modèles individuels dépendants des distances, dont une référence classique est Tomé and Burkhart (1989). Ces modèles sont particulièrement utilisés dans l'étude des processus biologiques (compétition entre espèces, étude de la mortalité et de la fertilité, etc).

Entre ces deux types de modèles, on peut distinguer d'autres types de modèles. Tout d'abord, les modèles de cohorte se situent entre les modèles individuels indépendants des distances et les modèles matriciels. Ce type de modèle repose sur la description d'une variable moyenne par classe (par exemple le diamètre moyen dans chaque classe) et où l'évolution de cette variable est régie par une fonction de croissance. Ce type de modèle a particulièrement été étudié par Alder (1979). On distingue ensuite les modèles de trouées ou « gap-model ». Une des caractéristiques de ces modèles est de modéliser les interactions entre les arbres indépendamment des distances à l'échelle d'une placette (Franc et al., 2000). Ils sont utilisés pour des populations sur des grandes superficies et pour des questions de succession ou d'évolution à long terme (face à des changements climatiques par exemple).

Modèles matriciels

Les modèles matriciels ont été introduits indépendamment par Bernardelli (1941), Lewis (1942) et particulièrement par Leslie (1945) en démographie humaine. Ces modèles reposent sur des matrices de transition décrivant les probabilités de passage d'un état à un autre, l'idée sous-jacente étant les chaînes de Markov. Le modèle le plus général est le modèle de Lefkovitch (1965) : toutes les transitions possibles entre états sont possibles, c'est-à-dire que la matrice stochastique comporte des éléments non nuls partout. On l'utilise quand les états sont des stades de développement, sans notion d'ordre, par exemple des états de végétation dans un paysage (Shugart et al., 1973 ; Horn, 1975). Le modèle de Usher (1966) suppose que les états peuvent être ordonnés. Il restreint les transitions possibles aux transitions d'un état dans lui-même ou dans l'état supérieur, c'est-à-dire que la matrice stochastique comporte des éléments non nuls sur sa diagonale et sur sa sous-diagonale. On l'utilise quand les états sont des classes de taille (population structurée en taille) et que la croissance est un processus du premier ordre (pas de saut

de classes, pas de régression). Le modèle de Leslie peut être vu comme un cas particulier du modèle de Usher : les seules transitions possibles sont d'un état dans l'état supérieur, c'est-à-dire que la matrice stochastique comporte des éléments non nuls sur sa sous-diagonale uniquement. On l'utilise quand les états sont des cohortes (ou classe d'âge). Une large présentation, ainsi que des développements et des résultats théoriques sont donnés par Caswell (2001).

Ces modèles sont beaucoup utilisés en biologie et particulièrement en écologie : lutte contre les espèces invasives (Neubert and Caswell, 2000), biologie de la conservation (Boyce, 1992 ; Menges, 2000 ; Alvarez-Buylla et al., 1996 ; Morris and Doak, 2002), gestion de populations, qu'elles soient animales (Hauser et al., 2006 ; Doubleday, 1975) ou végétales (Buongiorno and Gilless, 2003), etc. En foresterie, le modèle de Usher a été largement utilisé, par rapport au modèle de Lefkovitch par exemple qui autorise toutes les transitions. C'est un modèle bien adapté dans la modélisation forestière. En effet, la croissance des arbres ainsi que leur régénération est une évolution lente, comparée aux populations animales par exemple et dans l'échelle de temps d'étude d'une dizaine d'années. Ce travail de thèse se focalise sur le modèle de Usher et ses applications en foresterie. Une présentation ainsi que les résultats connus utilisés dans cette thèse sont donnés dans le premier chapitre.

Ces modèles sont utilisés dans la prédiction de grandeurs d'intérêt. Il est donc nécessaire de considérer le modèle matriciel comme un modèle stochastique. Nous verrons comment introduire de la stochasticité dans ces modèles.

Cas linéaire et densité-dépendant

Le modèles matriciels «classiques» reposent sur l'hypothèse que les évolutions des individus sont indépendantes entre elles et indépendantes du temps. Dans ce type de modèle, la relation, donnée au chapitre suivant par l'équation (1.1), et décrivant l'évolution de la population au cours du temps, est linéaire. Nous appellerons alors ce type de modèles des modèles matriciels linéaires. Leur comportement asymptotique est connu : ils conduisent à un état d'équilibre caractérisé par une croissance exponentielle de la population. Mais la croissance exponentielle à long terme de ces modèles est souvent peu réaliste. Un autre niveau de complexité est alors atteint lorsqu'on suppose que les évolutions des individus dépendent de la densité courante de la population. Ce sont les modèles densité-dépendants. La différence entre modèles matriciels linéaires et modèles matriciels densité-dépendants est du même ordre que celle entre le modèle de Malthus et le modèle logistique. Les modèles densité-dépendants rendent compte des phénomènes biologiques et

environnementaux de régulation dans l'évolution d'une population. En effet, lorsqu'une population croît, elle consomme ses ressources. Elle peut aussi être une proie pour les prédateurs ou être face à un phénomène de compétition interne. Enfin, son évolution est limitée par l'espace qu'elle occupe. Les relations de compétition entre individus dans un peuplement forestier est essentiellement due à l'accès des ressources qui lui sont nécessaires : lumière, eau, éléments minéraux. Ces phénomènes impliquent que le taux de survie et la fertilité sont réduites, et ces effets sont traduits par la densité-dépendance des paramètres d'évolution et de fécondité des individus (Franc et al., 2000 ; Caswell, 2001). Á l'inverse des modèles linéaires, les modèles densité-dépendants conduisent à différents types de comportement asymptotique : état stationnaire, cycles, pseudo-cycles et chaos. Ces comportements seront développés au chapitre suivant.

Stochasticité dans les modèles matriciels

On distingue trois types de stochasticité dans les modèles matriciels (Alvarez-Buylla et al., 1996) : démographique, environnementale et d'échantillonnage.

La stochasticité démographique (Caswell, 2001) consiste à décrire l'évolution de la population par une chaîne de Markov sur le vecteur d'effectif des classes d'états. Dans le modèle linéaire, comme les évolutions des individus sont indépendantes les unes des autres, cela revient à décrire une collection de trajectoires individuelles modélisées par des chaînes de Markov, indépendantes les unes des autres. Les flux d'individus provenant d'une classe donnée ne sont pas indépendants et leur loi conjointe est une loi multinomiale, conditionnelle au nombre d'individus issus de cette classe. Les effectifs recrutés, quant à eux, suivent une loi de Poisson. La stochasticité démographique a été intensément étudiée et est utilisée dans le cas de populations de petite taille pour évaluer le risque d'extinction (Boyce, 1992 ; Lande, 1993 ; Menges, 2000).

La stochasticité environnementale consiste à prendre en compte les effets environnementaux dans la variabilité des paramètres du modèle. Elle est intensément utilisée dans les études de viabilité de population (Boyce, 1992 ; Fieberg and Ellner, 2001 ; Kaye and Pyke, 2003). Les effets environnementaux peuvent être des effets spatiaux lié à l'hétérogénéité de la fertilité du sol, ou des effets temporels liés à des fluctuations temporelles, comme par exemple des changements climatiques.

Enfin, on peut décrire l'évolution de la population par un modèle déterministe. Les paramètres du modèle sont alors estimés à partir d'un échantillon, et ont une variabilité due à l'échantillonnage. Les estimateurs des paramètres de la matrice de transition sont alors des variables aléatoires, et les estima-

teurs des vecteurs d'effectif des vecteurs aléatoires. La variabilité d'échantillonnage est souvent négligée (Alvarez-Buylla et al., 1996 ; Taylor, 1995), alors que la variabilité environnementale et la variabilité démographique sont souvent prises en compte. En effet, la variabilité démographique est importante pour les petites populations pour prédire le risque d'extinction. De plus, quand on tient compte de la variabilité démographique ou de la variabilité environnementale, on apporte une information supplémentaire au modèle, contrairement à la variabilité d'échantillonnage qui peut apparaître au contraire comme une nuisance.

Objectifs de la thèse

Les forestiers utilisent des modèles de Usher pour prédire s'il y a équilibre entre les prélèvements et le renouvellement naturel de la ressource. Cette prédiction est entachée d'erreur car les paramètres du modèle sont estimés à partir d'échantillon de taille finie. La variabilité des prédictions des modèles matriciels est toutefois négligée. On souhaite donc apporter plus de rigueur dans l'utilisation des modèles de Usher en accompagnant chaque prédiction du modèle par son intervalle de confiance. La détermination de ces intervalles de confiance devront permettre de tester principalement si la population s'éteint, explose ou bien reste stationnaire. De plus, la précision de ces intervalles de confiance devra être reliée à la taille de l'échantillon, afin de déterminer le nombre d'arbres à inventorier pour avoir une précision donnée sur l'estimation des prédictions. Dans cette thèse, la stochasticité considérée (sauf mention explicite dans les cas contraires) sera ainsi la stochasticité d'échantillonnage.

On se placera, tout d'abord, dans le modèle linéaire. On s'intéressera au taux de croissance et à la distribution stationnaire de ce modèle. Ces prédictions seront vues comme des fonctions implicites des paramètres du modèle et seront estimées par maximisation de la vraisemblance. On déterminera alors le comportement asymptotique des estimateurs des prédictions, en utilisant l'efficacité du maximum de vraisemblance et la delta-méthode. Nous optons donc ici pour un point de vue asymptotique. Les résultats sur les prédictions sont ainsi applicables pour de grandes tailles d'échantillon. Nous préciserons alors par des simulations la taille d'échantillon à partir de laquelle cette méthode est satisfaisante.

On visera ensuite à affiner ces intervalles de confiance, en restant dans le modèle linéaire. Ceci revient à rechercher des estimateurs des paramètres de transition plus efficaces. Une façon d'améliorer les estimateurs des paramètres de transitions est de définir des estimateurs plus robustes. Cette amélioration est guidée par deux types de contraintes intrinsèques au modèle :

- la première contrainte est imposée par la structure discrète du modèle. Les erreurs de classification des individus influent sur la précision des estimations. Les estimateurs considérés devront alors être peu sensibles aux erreurs de classification que cette structure entraîne. Dans les modèles forestiers, ces erreurs proviennent en fait des erreurs de mesure sur les dispositifs permanents dont sont issues les données. Ces erreurs sont fréquentes, car les diamètres sont mal mesurés (erreurs d'imprécisions sur les mesures, arbres irréguliers, etc.). Utiliser des estimateurs robustes pour minimiser l'impact des erreurs de mesure est donc nécessaire ;
- la deuxième contrainte porte sur la dynamique de la population. Dans le modèle de Usher, le nombre de transitions entre les classes d'états est limité. Une fois le pas de temps et les classes d'état choisis, certaines données peuvent ne pas respecter cette contrainte. Nous proposerons un estimateur plus robuste que l'estimateur classique dans le cadre du traitement de ces données.

Nous nous placerons dans le modèle statistique continu, défini par la taille et l'accroissement en taille des individus à un pas de temps donné, et nous proposerons plusieurs façons de traiter ces questions, l'idée principale étant de tronquer l'échantillon. Nous formaliserons ensuite ces différents choix par différentes écritures des paramètres de transition. Les estimateurs de ces paramètres seront des estimateurs de la moyenne α-tronquée et plus généralement des L-estimateurs. Par ailleurs, nous utiliserons des critères de robustesse qui sont la sensibilité des estimateurs. Ils sont basées sur la notion de fonction d'influence. Nous calculerons alors la fonction d'influence des différents L-estimateurs dans le modèle statistique bidimensionnel.

Nous élargirons enfin les résultats du modèle linéaire au modèle de Usher densité-dépendant, où les évolutions des individus dépendent de la densité courante de la population. Dans ce travail, on considérera que les paramètres dépendent de façon exponentielle de la densité, et on s'assurera, dans un premier temps, qu'à une valeur des paramètres correspond un et un seul état stationnaire. Cette relation devra même être différentiable, afin d'utiliser des méthodes telles que la delta-méthode. Nous déterminerons enfin des intervalles de confiance du vecteur stationnaire.

Dans tout ce travail, les résultats théoriques seront appliqués à un jeu de données réelles sur un site expérimental à Paracou en Guyane Française.

Plan de la thèse

Cette thèse est organisée en quatre chapitres et des annexes. Le premier chapitre fait un état de l'art permettant de justifier des options adoptés

dans ce travail. Le deuxième et le troisième chapitre s'intéressent au modèle linéaire, et le quatrième chapitre au modèle densité-dépendant. Les annexes présentent trois articles et des compléments qui servent de base aux deuxième, troisième et quatrième chapitres.

Le deuxième chapitre est consacré à la détermination des intervalles de confiance des prédictions du modèle linéaire. Nous estimerons les paramètres du modèle par maximum de vraisemblance, et nous déterminerons le comportement asymptotique des estimateurs des prédictions ainsi obtenues.

Le troisième chapitre s'articule autour de la recherche d'estimateurs robustes au sens de la sensibilité. On cherchera des estimateurs robustes des paramètres de transition et on proposera un estimateur plus robuste que l'estimateur classique dans le cadre du traitement des données qui ne respecte pas la contrainte sur la dynamique imposée par les hypothèses du modèle de Usher.

Le quatrième chapitre étendra les résultats du modèle linéaire au cas du modèle densité-dépendant.

Dans un chapitre de conclusion, nous montrerons l'apport de ce travail et nous discuterons des limites. Nous proposerons alors des perspectives.

La première annexe est un article paru dans *Mathematical Biosciences* suivi de calculs complémentaires à cet article. Il sert de support au deuxième chapitre.

La deuxième annexe est un article sous presse dans *Computational Statistics and Data Analysis* et est un complément au troisième chapitre.

Enfin, la troisième annexe est un article soumis à *Acta Biotheoretica* et sert de base au quatrième chapitre.

Chapitre 1

État de l'art

Introduction

L'utilisation des modèles matriciels pour la gestion des ressources renouvelables forestières a été initiée par Usher (1966), qui a donné son nom au modèle de Usher. Ce modèle est particulièrement utilisé dans l'optimisation d'un scénario sylvicole (Michie and Buongiorno, 1984 ; Vanclay, 1995 ; Favrichon and Young Cheol, 1998), dans des études économiques (Lu and Buongiorno, 1993) ou climatiques (Boscolo et al., 1999). Il a été largement utilisé en foresterie, par rapport aux autres modèles matriciels et en particulier par rapport au modèle de Lefkovitch. Certains travaux ont utilisés ce modèle (Solomon, 1986 ; Lamar and McGraw, 2005), mais celui-ci n'apportait pas beaucoup plus d'information : les classes de diamètres et le pas de temps étant bien choisis, peu de données avaient une évolution différente de celle décrite par le modèle de Usher.

Les modèles matriciels sont des modèles agrégés, c'est-à-dire qu'ils donnent une description résumée des dynamiques individuelles par une description de l'évolution de la population dans chaque classe d'état. Cette évolution est reliée à celle des individus par les composantes du modèles qui sont la croissance, la régénération et la mortalité des individus de chaque classe d'état. Ce type de modèles est particulièrement intéressant dans l'étude d'un peuplement forestier tropical et de grande taille (Favrichon, 1998). Ce type de peuplement a en effet pour particularité un développement et une structure hétérogène : il est composé d'une grande diversité d'espèces, et qui ont des comportements et des potentiels différents dans la lutte pour la survie. L'utilisation des modèles matriciels ne rend pas compte de la diversité de ces comportements mais permet de simplifier la dynamique du peuplement. Ce type de modèle est bien adapté pour des échantillons de grandes tailles, mais

en pratique on l'utilise aussi pour des peuplements de petites tailles dans le cadre d'aménagement forestier. Il permet dans ce cas de tester la viabilité d'un peuplement de façon rapide et simple (Boyce, 1992 ; Menges, 2000 ; Alvarez-Buylla et al., 1996 ; Morris and Doak, 2002).

Dans cette thèse, nous nous intéressons au modèle de Usher pour son intérêt dans l'étude d'un peuplement forestier.

1.1 Définition du modèle de Usher

1.1.1 Modèle de Usher linéaire

Le modèle de Usher (Usher, 1966, 1969) est un modèle matriciel qui décrit l'évolution en temps discret d'une population structurée par taille. Il est basé sur quatre hypothèses (Favrichon, 1998) :

- hypothèse d'indépendance : les évolutions des individus sont indépendantes ;
- hypothèse markovienne : l'évolution d'un individu entre deux pas de temps t et $t+1$ dépend de son état au temps t ;
- hypothèse de Usher : pendant un pas de temps, un individu peut soit rester dans sa classe, soit passer dans la classe supérieure, ou soit mourir ; chaque individu peut donner naissance à de nouveaux individus dans la première classe ;
- hypothèse de stationnarité : les évolutions des individus sont indépendantes du temps.

Ces quatres hypothèses constituent les hypothèses du modèle de Usher *stricto sensu*, mais un certain nombre de travaux ont visé à supprimer en partie ces hypothèses. Celles-ci sont en effet des simplifications du comportement d'une population et sont discutables. Ceci est en particulier vrai dans le cas d'un peuplement forestier (Picard et al., 2003). Par extension, on parle toujours de modèle de Usher même quand certaines de ces hypothèses ont été supprimées. Passons à présent en revue chacune de ces hypothèses en voyant comment les supprimer.

L'hypothèse de Usher impose une contrainte sur la dynamique de la population, qui n'existe pas dans le modèle de Lefkovitch par exemple qui autorise toutes les transitions. Néanmoins, le modèle de Usher est bien adapté dans l'étude de la dynamique forestière. Á condition de bien choisir le pas de temps et les classes de diamètre, l'hypothèse de Usher est une hypothèse réaliste. De plus, ce modèle est avantageux car la récolte de données est souvent très coûteuse en temps et les données longitudinales sont donc limitées. Ainsi, la réduction du nombre de paramètres dans la matrice de transition

permet d'obtenir des estimateurs de ces paramètres plus précis. Dans cette thèse, on s'intéressera en particulier à l'hypothèse de Usher *via* la définition d'un nouvel estimateur robuste des taux de transition (voir chapitre 3). En ce qui concerne l'hypothèse d'indépendance, elle est fausse biologiquement car les arbres interagissent entre eux (pour la compétition pour la lumière par exemple) et leur évolution dépend de leur localisation spatiale. Cependant, d'un point de vue mathématique, on arrive à prédire correctement la dynamique de la population en faisant l'hypothèse d'indépendance (au sens statistique du terme) entre les arbres. La formulation du problème biologique de compétition entre arbres en un problème mathématique a conduit à une simplification, qui s'avère en pratique justifiée. Quand à l'hypothèse de Markov, elle n'est pas réaliste par rapport aux phénomènes biologiques, puisqu'un arbre qui a bien poussé une année continue à bien pousser l'année suivante, et inversement un arbre qui pousse lentement a beaucoup plus de chance de mourir (Horn, 1975 ; van Hulst, 1980 ; Picard et al., 2003 ; Johnson et al., 1991). Enfin, l'hypothèse de stationnarité est supprimée dans les modèles densité-dépendants (Favrichon, 1998 ; Boscolo and Vincent, 1998). Ce type de modèle est développé plus loin.

Formellement, on suppose qu'il y a I classes d'état et que, le temps étant discret, il est indexé par un entier t et que la durée d'un pas de temps est notée Δt. L'évolution de la population est décrite par le vecteur d'effectif, noté $N(t) = (N_i(t))_{i=1,\dots,I}$, et définie entre le temps t et $t+1$ par la relation :

$$N(t+1) = U\,N(t) \tag{1.1}$$

où U est la matrice de transition définie par :

$$U = \begin{pmatrix} p_1 + f_1 & f_2 & \cdots & f_I \\ q_2 & p_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & q_I & p_I \end{pmatrix}$$

où p_i est le taux de transition des individus de la classe i qui restent vivants dans leur classe, q_{i+1} celui des individus qui restent vivants et passent dans la classe supérieure, et f_i le taux de fécondité de la classe i. Le taux de mortalité des individus de la classe i est $m_i = 1 - p_i - q_{i+1}$. Ceci implique que $p_i \geq 0$, $q_{i+1} \geq 0$ et $p_i + q_{i+1} \leq 1$. On suppose de plus que $p_i > 0$ (sinon, le nombre de classes I peut être réduit), $q_i > 0$ et $f_i > 0$. Dans certaines situations, il n'est pas possible d'estimer un taux de fécondité par classe, mais seulement un taux moyen de fécondité f avec : $f_1 = \ldots = f_I = f$.

Alternativement, la matrice de Usher peut s'écrire $U = PS + R$, où :

$$P = \begin{pmatrix} 1-q_1^\bullet & & 0 & 0 \\ q_1^\bullet & \ddots & & \vdots \\ & \ddots & 1-q_{I-1}^\bullet & 0 \\ 0 & & q_{I-1}^\bullet & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 1-m_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & 1-m_I \end{pmatrix}$$

$$R = \begin{pmatrix} f_1 & \dots & f_I \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix}$$

et où $q_i^\bullet$ est le taux de transition conditionnel des individus de la classe i dans la classe $i+1$ sachant qu'ils restent vivants. Ainsi la matrice P décrit la croissance de la population, S est la matrice de survie et R décrit le recrutement. De plus, on a la relation :

$$q_i^\bullet = \frac{q_{i+1}}{1-m_i}$$

Le modèle de Usher repose en particulier sur l'hypothèse d'indépendance et de stationnarité. Ces hypothèses peuvent être relâchées en considérant le modèle de Usher densité-dépendant.

1.1.2 Modèle de Usher densité-dépendant

Le modèle densité-dépendant est une extension du modèle linéaire. L'hypothèse d'indépendance et de stationnarité ne sont plus valables. Au temps t, ce modèle s'écrit (Caswell, 2001) :

$$N(t+1) = U_t\, N(t)$$

La densité-dépendance signifie aussi que la matrice de Usher U au temps t (ou de façon équivalente, les taux de transition p_i, q_i et f_i) dépendent du vecteur $N(t)$, c'est-à-dire :

$$U_t = U(N(t))$$

Les paramètres du modèle densité-dépendant sont traditionnellement classés suivant qu'ils sont compensateurs, sur-compensateurs et décompensateurs. Cette classification provient du modèle unidimensionnel, dans le cas de populations non structurées, et qui s'écrit de la façon suivante :

$$\begin{aligned} N(t+1) &= u(N(t))\, N(t) \\ &= v(N(t)) \end{aligned}$$

Le taux de croissance u est dit décompensateur si c'est une fonction croissante de N, c'est-à-dire, si pour tout $N > 0$:

$$\frac{d\,u(N)}{d\,N} > 0$$

Ce type de modèles peut être utilisé pour des populations végétales, où une densité limitée réduit la fertilité.

Si $\frac{d\,u(N)}{d\,N} \leq 0$, et que :

$$\begin{aligned} \frac{d\,v(N)}{d\,N} &\geq 0 \\ \lim_{N\to\infty} v(N) &= L \end{aligned}$$

où L est une constante strictement positive, u est dit compensateur. La décroissance de u comme fonction de N compense exactement la croissance de la densité.

Si $\frac{d\,u(N)}{d\,N} \leq 0$ et que :

$$\lim_{N\to\infty} v(N) = 0$$

u est dit sur-compensateur. La croissance de la densité n'est pas assez importante pour compenser la décroissance de u.

Les deux fonctions du taux de croissance les plus connues sont celles de Beverton-Holt (équation (1.2)) et de Ricker (équation (1.3)) :

$$u(N) = \frac{1}{1 + c\,N} \tag{1.2}$$

$$u(N) = e^{-c\,N} \tag{1.3}$$

où c est une constante. La fonction de Beverton-Holt est compensatrice, et celle de Ricker est sur-compensatrice.

Dans les modèles matriciels, la définition de compensateur, sur-compensateur et décompensateur s'applique pour chaque paramètre de la matrice U. Les paramètres de transition doivent de plus vérifier la contrainte de sous-stochasticité, c'est-dire que la somme des taux de transition d'une même classe doit être inférieure à un.

Par ailleurs, une approximation souvent utilisée est d'écrire les taux de transition comme une fonction du produit scalaire $N(t) \cdot a$, où $\cdot$ désigne le produit scalaire et a est un vecteur de $\mathbb{R}^I$. Un cas particulier est de considérer le vecteur a dont les coordonnées sont égales à 1. Les paramètres sont alors des fonctions de l'effectif total. La quantité $N(t) \cdot a$ est de plus souvent appelée l'indice de compétition. En effet, si les taux de transition sont des

fonctions de Berverton-Holt ou de Ricker, leur expression est celle des équations respectivement (1.2) et (1.3) en remplaçant N par $N(t) \cdot a$. Ainsi, si $c > 0$, la dérivée des taux de transition par rapport à $N(t) \cdot a$ est négative.

1.2 Justification de l'utilisation du modèle de Usher

L'objectif appliqué est que les acteurs de la gestion forestière utilisent des modèles matriciels. Ceux-ci utilisent des formules telles que la formule de reconstitution du stock (Durrieu de Madron et al., 1998), noté $\tilde{R}$. Celui-ci est défini à partir du stock exploitable, S, qui correspond à l'effectif total d'individus à partir d'une classe donnée J (correspondant au diamètre minimum d'exploitabilité dans le cas d'un aménagement forestier) et de la durée de rotation, T. Il s'écrit comme le ratio du stock exploitable au temps T sur le stock exploitable au temps 0, sachant qu'entre les instants 0 et 1 la population a subi une exploitation qui a enlevé tous les individus dans la classe donnée et les classes supérieures, soit :

$$\tilde{R} = \frac{S(T)}{S(0)}$$

Le stock exploitable à l'instant t, $S(t)$, s'écrit sous la forme :

$$S(t) = \sum_{i=J}^{I} N_i(t) = \widetilde{C} N(t)$$

où $\widetilde{C}$ est le vecteur transposé du vecteur C de longueur I dont les $J-1$ premiers éléments sont nuls et les $(I-J+1)$ suivant sont égaux à un. A l'instant $t = 0$, les arbres exploitables sont coupés. On note $N(0^+)$ le vecteur d'effectifs juste après l'exploitation. Il s'écrit : $N(0^+) = N(0) \odot (\mathbf{1} - C)$ où $\mathbf{1}$ est le vecteur de longueur I dont les coordonnées sont égales à un et où $\odot$ désigne le produit terme à terme. Par récurrence, on obtient l'expression de $S(t)$ suivante :

$$S(t) = \widetilde{C} U^t N(0^+)$$

$\tilde{R}$ s'écrit alors en fonction de la matrice de Usher U comme suit :

$$\tilde{R} = \frac{\widetilde{C} U^T N(0^+)}{\widetilde{C} N(0^+)}$$

Les modèles de Usher sont ainsi une extension des outils mathématiques utilisés par les gestionnaires forestiers. Les modèles proposés doivent de plus

rester *simples*. Les modèles de Usher concilient simplicité et adéquation aux besoins.

De plus, les modèles individus-centrés peuvent être agrégés en modèle arbre indépendant des distances et ceux-ci sont équivalents, sous des conditions assez peu contraignantes, à des modèles de distribution. La distinction entre distribution continue (modèle à base d'équations aux dérivées partielles) et distribution discrète (modèles matriciels) n'a pas lieu d'être en pratique (ce point sera justifié dans la section suivante). Le modèle de Usher émerge donc naturellement comme le résultat de l'agrégation de modèles plus complexes.

D'autre part, les prédictions faites avec les modèles de Usher s'avèrent de qualité comparable à celles faites avec des modèles plus complexes (Picard et al., 2004). Le rasoir d'Occam (ou principe de parcimonie) suggère dans ce cas de s'en tenir au modèle le plus simple.

1.3 Propriétés du modèle de Usher

Quelques unes des propriétés du modèle de Usher, importantes pour ce travail de thèse, sont présentées ici.

Le modèle matriciel peut être tout d'abord vu comme la limite d'une collection de chaînes de Markov indépendantes d'une part, et comme la limite d'un schéma numérique d'intégration d'une équation aux dérivées partielles d'autre part. Ainsi le modèle matriciel est à l'intersection de plusieurs voies de modélisation.

De plus, le modèle matriciel a un comportement asymptotique. Il est donc naturel dans un premier temps de s'intéresser à ce comportement asymptotique comme prédiction du modèle. L'objectif de la thèse sera donc de construire des intervalles de confiance pour les grandeurs caractérisant l'état asymptotique du modèle.

1.3.1 Lien entre le modèle déterministe et les chaînes de Markov

Si on suppose que l'évolution de la population est décrite de façon stochastique selon le point de vue de la stochasticité démographique, le vecteur d'effectif au temps t, $N(t) = (N_i(t))_{i=1\dots I}$, est un vecteur aléatoire de $\mathbb{N}^I$. L'évolution du vecteur d'effectif entre le temps t et $t+1$ est donnée par la relation :

$$N_i(t+1) = F_{ii}(t) + F_{i-1i}(t) \tag{1.4}$$

où $F_{ij}(t)$ est le flux d'individus de la classe i à la classe j entre les instants t et $t+1$ (avec pour convention $F_{01}(t) = r(t)$, où $r(t)$ est le nombre de nouveaux

individus). On note également $F_{i\dagger}(t)$ le nombre d'individus de la classe i qui meurent entre t et $t+1$. La loi des flux conditionnellement à $N(t)$ est une loi multinomiale, et le nombre de recrutés est tiré suivant une loi de Poisson :

$$\begin{aligned}(F_{ii}(t), F_{ii+1}(t), F_{i\dagger}(t)) &\sim \mathcal{M}\left(N_i(t), p_i(N(t)), q_{i+1}(N(t)), 1 - p_i(N(t)) - q_{i+1}(N(t))\right)\\ r(t) &\sim \mathcal{P}(\sum_{i=1}^{I} f_i(N(t))N_i(t))\end{aligned}$$

La suite $(N(t))_{t\in\mathbb{N}}$ est une chaîne de Markov de $\mathbb{N}^I$ de probabilité de transition notée Π définie, pour tout $m = (m_1, \ldots, m_I)$ et $n = (n_1, \ldots, n_I)$ de $\mathbb{N}^I$, par :

$$\Pi\left((n_1, \ldots, n_I), (m_1, \ldots, m_I)\right) = \prod_{i=1}^{I} \Pi_i(n_i, m)$$

où $\Pi_i(n_i, m) = \Pr\left(N_i(t+1) = n_i | N(t) = m\right)$ est, pour $i \geq 2$, la convolution de deux binomiales, c'est-à-dire :

$$\Pi_i(n_i, m) = \sum_{k=n_i-\min(m_{i-1},n_i)}^{\min(m_i,n_i)} C_{m_i}^{k} p_i(m)^k (1-p_i(m))^{m_i-k} C_{m_{i-1}}^{n_i-k} q_i(m)^{n_i-k} (1-p_i(m))^{m_{i-1}-n_i+k}$$

si $n_i \leq m_i + m_{i-1}$ et $\Pi_i(n_i, m) = 0$ sinon. Pour $i = 1$, $\Pi_1(n_1, m)$ est la convolution d'une loi binomiale et d'une loi de Poisson :

$$\Pi_1(n_1, m) = \sum_{k=0}^{\min(m_1,n_1)} C_{m_1}^{k} p_1(m)^k (1-p_1(m))^{m_1-k} \frac{\left(\sum_{j=1}^{I} f_j(m) m_j\right)^{n_1-k}}{(n_1-k)!} e^{-\sum_{i=j}^{I} f_j(m) m_j}$$

De plus, l'équation (1.4) donne l'évolution du vecteur d'effectif par la relation :

$$\mathbb{E}[N(t+1)|N(t)] = U(N(t))\, N(t) \tag{1.5}$$

où U est la matrice de Usher. Dans le cas du modèle linéaire, cette équation s'écrit :

$$\mathbb{E}[N(t+1)] = U\, \mathbb{E}[N(t)] \tag{1.6}$$

Ainsi, le modèle déterministe est la version en moyenne du modèle stochastique, démographiquement parlant.

De plus, dans le modèle linéaire, les évolutions des individus sont indépendantes entre elles. L'évolution de la population peut alors être décrite par une collection de trajectoires individuelles modélisées par des chaînes de Markov, dans un espace d'états fini, indépendantes les unes des autres. Si $((X_1^t, X_1^{t+1}), \ldots, (X_n^t, X_n^{t+1}))$ est l'échantillon qui décrit l'état des individus

au temps t et $t+1$, les vecteurs d'effectif de la classe C_i aux temps t et $t+1$ s'écrivent :

$$N_i(t) = \sum_{k=1}^{n} \mathbf{1}_{X_k^t \in C_i}, \qquad N_i(t+1) = \sum_{k=1}^{n} \mathbf{1}_{X_k^{t+1} \in C_i}$$

La loi faible des grands nombres pour des variables aléatoires bornées dans L^1 donne le comportement asymptotique de $N_i(t)$ et $N_i(t+1)$, pour tout $i = 1, \ldots, I$:

$$\begin{aligned} \frac{N_i(t)}{n} &\xrightarrow[n\to\infty]{L^1} y_i(t) \\ \frac{N_i(t+1)}{n} &\xrightarrow[n\to\infty]{L^1} y_i(t+1) \end{aligned}$$

où $y(t) = (y_i(t))_{i=1,\ldots,I}$ est le vecteur densité de $[0,1]^I$, dont les composantes sont égales à : $y_i(t) = \Pr(X^t \in C_i)$. On a la même définition pour $y(t+1) = (y_i(t+1))_{i=1,\ldots,I}$ au temps $t+1$. Par la relation (1.6), on a de plus :

$$\mathbb{E}\left[\frac{N(t+1)}{n}\right] \xrightarrow[n\to\infty]{} U\, y(t)$$

On en déduit que $y(t+1) = U\, y(t)$. La relation (1.6) divisée par la taille de la population, n, converge alors, lorsque n tend vers l'infini, vers le modèle déterministe, vérifié par le vecteur densité de la population y. Le modèle déterministe linéaire est donc une approximation grande population du modèle stochastique, démographiquement parlant. C'est de plus un modèle qui décrit en fait l'évolution de la densité de la population dans chaque classe. Lorsqu'on utilise le modèle matriciel de Usher, on fait l'approximation que le vecteur d'effectif vérifie la relation de récurrence traduisant l'évolution de la population, en supposant que c'est un vecteur de $\mathbb{R}_+^I$. De plus, par la relation de Chapman-Kolmogorov, on en déduit que les taux de transition de la matrice de Usher sont en fait les probabilités de transition des chaînes de Markov individuelles.

1.3.2 Lien entre le modèle matriciel et les équations aux dérivées partielles

Le modèle matriciel peut également être vu comme une approximation d'un système dynamique en temps et taille continus basé sur une équation aux dérivées partielles (EDP). Plus précisément, la limite lorsque la largeur des classes et le pas de temps d'un modèle matriciel tendent vers zéro est

un système dynamique à base d'EDP (au sens d'une limite appropriée). Une preuve de ces limites pour des modèles matriciels de Leslie figure dans Keyfitz (1967) et Goodman (1967).

Nous donnons ici un aperçu de ce lien dans le sens inverse, à savoir : le modèle matriciel correspond à un schéma numérique d'intégration d'une EDP du premier ordre. Prenons l'exemple de l'équation de Liouville avec puits :

$$\frac{\partial \phi}{\partial t} = -\frac{\partial (a\phi)}{\partial x} - m\phi \tag{1.7}$$

avec la condition aux limites :

$$a(x_0)\,\phi(x_0, t) = r$$

Dans cet exemple, ϕ représente la distribution diamétrique (continue) au temps t, c'est-à-dire : $\phi(x,t)\,dx$ est le nombre d'arbres de diamètre compris entre x et $x+dx$ au temps t. Le temps (t) et le diamètre (x) sont continus. La fonction $x \mapsto a(x)$ représente la vitesse de croissance en diamètre, tandis que $x \mapsto m(x)$ désigne le taux de mortalité d'un arbre de diamètre x. r désigne le flux de recrutement. L'équation de Liouville est une équation classique pour décrire le transport (Gardiner, 1985). Le transport en question correspond ici à la croissance des arbres. Un terme de puits, $-m\phi$, a été rajouté pour décrire la mortalité des arbres. La condition aux limites traduit le recrutement. Cette équation de Liouville avec puits peut également être vue comme l'équation maîtresse d'un processus stochastique (Suzuki and Umemura, 1974).

L'EDP (1.7) admet une solution analytique relativement simple. Envisageons cependant de la résoudre de façon numérique. L'espace des diamètres est discrétisé en petits intervalles $i\delta$, $i \in I \subset \mathbb{N}$; le temps est de même discrétisé en petits intervalles $k\tau$, $k \in K \subset \mathbb{N}$. Soit ϕ_{ik} la valeur de ϕ au point ($i\delta$, $k\tau$) et ϕ_k le vecteur $(\phi_{1k}, ..., \phi_{Ik})$. On note de même a_i la valeur de a au point $i\delta$ et m_i la valeur de m au point $k\tau$. Un schéma numérique explicite décentré à gauche de (1.7) s'écrit alors (Press et al., 1992) :

$$\begin{aligned} \frac{\phi_{ik+1} - \phi_{ik}}{\tau} &= -\frac{a_i\phi_{ik} - a_{i-1}\phi_{i-1k}}{\delta} - m_i\phi_{ik} \quad (i > 0) \\ a_0\phi_{0k+1} &= r \end{aligned}$$

Ce schéma peut être réécrit sous la forme matricielle :

$$\phi_{k+1} = \mathbf{U}\,\phi_k + \mathbf{r} \tag{1.8}$$

où la matrice $\mathbf{U}$ comporte les éléments $1 - a_i\tau/\delta - m_i\tau$ sur sa diagonale, les éléments $a_{i-1}\tau/\delta$ sur sa sous-diagonale, et des zéros partout ailleurs ; le

vecteur **r** comporte r en premier élément et des zéros partout ailleurs. **U** a la forme d'une matrice de Usher et l'équation (1.8) s'identifie à un modèle de Usher. Ce calcul simple s'étend à des systèmes non autonomes (on obtient alors un modèle de Usher non stationnaire).

Le succès d'un schéma numérique pour intégrer une EDP dépend des pas δ et τ choisis pour discrétiser l'espace des phases et le temps. De la même manière, un modèle de Usher pourra être vu comme une bonne approximation d'un système dynamique à base d'EDP à condition que son pas de temps et ses largeurs de classes soient correctement choisis. L'analogie va encore plus loin puisque la forme de l'élément $a_{i-1}\tau/\delta$ qui figure sur la sous-diagonale de la matrice **U** du schéma numérique a été à la base de la définition d'un estimateur des taux de transition pour les modèles de Usher (cf. section 1.4).

Le fait que le modèle matriciel peut être vu comme une approximation d'un système dynamique dans un espace continu justifie que l'on se focalise dessus.

1.3.3 Comportement asymptotique du modèle

Modèle de Usher linéaire

Le comportement asymptotique du modèle de Usher est déterminé à l'aide du théorème de Perron-Frobénius (Caswell, 2001). C'est un théorème algébrique qui décrit le spectre des matrices positives (Gantmacher, 1959). On suppose ici que U est une matrice irréductible, apériodique et on démontre que, dans ce cas, le modèle tend vers un état d'équilibre, caractérisé par la première valeur propre de U et son vecteur propre associé. Pour cela, on cherche les solutions de l'équation :

$$N(t+1) = U\,N(t)$$

Ceci donne, pour tout t :

$$N(t) = U^t\,N(0)$$

Dans le cas d'une matrice irréductible, apériodique, le théorème de Perron-Frobénius assure qu'il existe une et une seule valeur propre de U réelle, strictement positive, λ_1, telle que pour toute autre valeur propre de U dans $\mathbb{C}$, $\widetilde{\lambda}$, $\lambda_1 > |\widetilde{\lambda}|$. De plus, le vecteur propre associé w_1, de norme égale à un, est de composantes strictement positives. On note λ_k, pour $k = 2\ldots, I$, les autres valeurs propres de U, rangées dans l'ordre décroissant en module. Les vecteurs propres associés et de norme un sont notés w_k. Dans la base

$(w_1, \ldots, w_I)$, $N(0)$ s'écrit :

$$N(0) = \sum_{k=1}^{I} c_k \, w_k$$

les coefficients c_k étant des complexes. Ainsi, en utilisant la relation $U \, w_k = \lambda_k \, w_k$, on obtient :

$$N(t) = \sum_{k=1}^{I} c_k \, \lambda_k^t \, w_k \tag{1.9}$$

L'équation (1.9) donne :

$$\frac{N(t)}{\lambda_1^t} = c_1 \, w_1 + \sum_{k=2}^{I} c_k \, w_k \, \left(\frac{\lambda_k}{\lambda_1}\right)^t \tag{1.10}$$

On en déduit que :

$$\lim_{t \to \infty} \frac{N(t)}{\lambda_1^t} = c_1 \, w_1$$

$N(t)$ tend donc vers un état d'équilibre, où λ_1 représente le taux de croissance, et w_1 sa répartition suivant les classes d'état, et ceci indépendamment de $N(0)$.

De plus, la deuxième valeur propre détermine la vitesse de convergence vers cet état d'équilibre (Caswell, 2001). En effet, soit ρ défini par : $\rho = \frac{\lambda_1}{|\lambda_2|}$. D'après l'équation (1.10), on obtient :

$$\lim_{t \to \infty} \rho^t \, \left(\frac{N(t)}{\lambda_1^t} - c_1 \, w_1\right) = c_2 \, w_2$$

On en déduit donc qu'il existe K, tel que pour t assez grand,

$$\left| \frac{N(t)}{\lambda_1^t} - c_1 \, w_1 \right| \leq K \, \rho^{-t}.$$

La vitesse de convergence vers l'état d'équilibre est donc exponentielle, avec un taux d'au moins $\log(\rho)$.

Modèle de Usher densité-dépendant

Contrairement au modèle linéaire, le modèle densité-dépendant conduit à différents types de comportement asymptotique :

– état stationnaire : il existe un vecteur N de $\mathbb{R}^I$, appelé point d'équilibre qui vérifie l'équation :

$$N = U(N) \, N$$

– cycles : un cycle de période k est la donnée d'un ensemble $\mathcal{S} = \{n_1, n_2, \ldots, n_k\}$ tel que si $N(t) = n_i \in \mathcal{S}$, alors :

$$N(t+k) = n_i \qquad \text{et} \qquad N(t+l) \neq n_i \quad \text{pour } 0 < l < k$$

– pseudo-cycles : la dynamique du modèle consiste en des oscillations.
– chaos : le modèle ne converge vers aucun des comportements précédents.

Ces différents types de comportement dépendent de la forme des paramètres comme des fonctions du vecteur d'effectif courant et de la valeur des constantes qui déterminent ces fonctions. Suivant la valeur que l'on donne à ces constantes au départ, le modèle peut aboutir à n'importe lequel de ces comportements.

D'autre part, dans le cas stationnaire, il n'y a pas forcément unicité de l'état stationnaire. A une valeur des paramètres de la matrice de transition peut correspondre plusieurs points d'équilibre. De plus, une petite modification des paramètres autour d'une valeur donnée peut entraîner un changement dans l'état stationnaire. L'expression algébrique du point d'équilibre en fonction des paramètres n'est alors pas différentiable en cette valeur. Ceci empêche par exemple tout raisonnement basé sur la méthode delta car elle consiste en un développement de Taylor à l'ordre 1. Elle permet de relier la variabilité des prédictions du modèle à la variabilité des estimateurs des paramètres.

De façon générale, on appelle bifurcation un changement dans la dynamique asymptotique du modèle causé par une petite modification des paramètres de la matrice de transition. Lorsqu'une bifurcation apparaît, la dynamique du modèle est alors instable. En pratique, pour étudier les bifurcations on utilise un diagramme de bifurcation. Il consiste en des simulations, où l'idée est de fixer tous les paramètres sauf un, et de simuler l'état asymptotique du modèle pour chaque valeur de ce paramètre.

Cushing (1988) ; Cushing and Zhou (1994), dans ses travaux, s'est intéressé à la convergence des modèles vers un état stationnaire, en s'appuyant sur la théorie des bifurcations. Il s'avère judicieux de travailler sur le taux de reproduction net n, qui représente le nombre moyen de descendants qu'un individu a au cours de sa vie, plutôt que sur le paramètre malthusien λ. Un résultat établi par Cushing est que le signe de $\lambda - 1$ est le même que le signe de $n - 1$. L'intérêt de cette grandeur n est que l'information pour l'estimer est en général disponible et que son expression algébrique est souvent différentiable.

1.4 Les différents estimateurs des taux de transition

1.4.1 Estimateurs dans le modèle de Usher linéaire

Tout en conservant les caractéristiques et la complexité du modèle mathématique, en particulier sa structure discrète et l'hypothèse de Usher, la recherche d'estimateurs des paramètres de transition s'est faite dans un souci d'obtenir des estimateurs les plus précis possibles, et qui utilisent au mieux l'information disponible. On peut alors construire des estimateurs des paramètres de transition dans :

- le modèle statistique discret : les variables aléatoires sont les évolutions en classes d'états des individus.
- le modèle statistique continu : les variables aléatoires sont les évolutions en taille des individus.

Estimateurs du modèle discret

Dans le modèle statistique discret, les variables aléatoires sont les flux de populations de la classe i à la classe j entre t et $t+1$, notés $F_{ij}(t)$, pour $i = 1, \ldots, I$ et $j = i, i+1, \dagger$. Michie and Buongiorno (1984) ont comparé quatre méthodes en se basant sur les équations intrinsèques du modèle de Usher. Tout d'abord, la relation entre les paramètres de transition d'une même classe i s'écrit :

$$m_i = 1 - p_i - q_{i+1} \tag{1.11}$$

De plus, le modèle de Usher, entre le temps t et $t+1$, peut s'écrire sous forme développée :

$$\begin{cases} N_1(t+1) &= p_1 N_1(t) + \sum_{k=1}^{I} f_i N_i(t) \\ N_i(t+1) &= p_i N_i(t) + q_i N_{i-1}(t) \qquad (i > 1) \end{cases} \tag{1.12}$$

ou encore :

$$\begin{cases} N_I(t+1) - (1-m_I) N_I(t) &= q_I N_{I-1}(t) \\ N_i(t+1) - (1-m_i-q_{i+1}) N_i(t) &= q_i N_{i-1}(t) \qquad (i = 2 \ldots I-1) \\ N_1(t+1) - (1-m_1-q_2) N_1(t) &= \sum_{k=1}^{I} f_i N_i(t) \end{cases} \tag{1.13}$$

De ces équations, ils en tirent quatre estimateurs :

- les estimateurs du maximum de vraisemblance d'une loi multinomiale, ce qui revient à estimer les probabilités q_{i+1} et m_i par les proportions F_{ii+1}/N_i et $F_{i\dagger}/N_i$;
- les estimateurs de régression linéaire : on réalise I régressions linéaires indépendantes de $N_i(t+1)$ par rapport à $N_i(t)$ et $N_{i-1}(t)$, selon les équations (1.12) ;
- les estimateurs du modèle linéaire généralisé : on estime d'un seul coup par le modèle linéaire généralisé tous les coefficients de (1.12), en tenant compte des corrélations données par les relations (1.11) ;
- les estimateurs mixte de proportion et modèle linéaire : on estime les probabilités m_i par la proportion $F_{i\dagger}/N_i$. La probabilité q_I est ensuite estimée par régression linéaire selon (1.13), et on estime alors q_{I-1} par régression linéaire. On réitère ainsi le procédé en estimant les paramètres en cascade jusqu'à q_2.

Michie and Buongiorno (1984) ont conclu que l'estimateur du maximum de vraisemblance issu de la première méthode était meilleur car c'est le seul estimateur non biaisé parmi les estimateurs proposés.

Une autre méthode (Houde and Ledoux, 1995 ; Ingram and Buongiorno, 1996) consiste à estimer les probabilités de transition d'une matrice de Usher en supposant le peuplement dans l'état stationnaire et connaissant la distribution en taille stationnaire. Elle utilise le fait que la distribution en taille, le taux de mortalité et la vitesse de croissance en taille sont liés à l'état stationnaire par la relation $N = UN$. Cette méthode est utilisée pour des forêts non perturbées et dont les structures floristiques et diamétriques n'évoluent que très faiblement par rapport à l'échelle de temps étudié.

Dans le cadre bayésien, les coefficients d'une matrice de Usher peuvent également être estimés en utilisant la relation (Rivot, 2003) :

$$[\theta|F] = \frac{[\theta][F|\theta]}{\int [\theta][F|\theta]}$$

où θ désigne l'ensemble des coefficients de la matrice, F désigne l'ensemble des flux F_{ii}, F_{ii+1} et $F_{i\dagger}$, $[\theta]$ est la distribution a priori de θ, $[\theta|F]$ est la distribution a posteriori de θ, et $[F|\theta]$ est la distribution conditionnelle de F sachant θ. L'estimation consiste en fait à calculer la distribution a posteriori des coefficients. L'approche bayésienne a surtout un intérêt dans un contexte hiérarchique, quand il faut tenir compte de probabilités d'observation des individus pour estimer les taux de transition. Elle prend tout son sens en biologie animale, avec les problématiques de capture–recapture, mais n'a que peu d'intérêt dans un contexte forestier où les individus sont observés avec certitude.

Dans le modèle statistique discret, l'estimateur du maximum de vraisemblance est non biaisé. Dans le modèle statistique continu, cet estimateur défini par proportion ne correspond plus à l'estimateur du maximum de vraisemblance et reste non biaisé. Néanmoins, cet estimateur est très instable dans le cas de petits échantillons (Rogers-Bennett and Rogers, 2006). Nous le verrons également dans cette thèse dans le chapitre 3. En fait, cet estimateur n'exploite pas l'information contenue dans la taille.

Estimateurs du modèle continu

Gross et al. (2006) proposent une méthode pour affiner les estimateurs des paramètres afin d'éviter une sur-estimation lorsque les données sont limitées. Ils utilisent l'estimateur du maximum de vraisemblance du taux de survie s relié à l'information continue, en supposant que s est l'inverse d'une fonction logit d'une combinaison linéaire de la taille x comme suit :

$$s(x, a_0, a_1) = \frac{\exp(a_0 + a_1 x)}{1 + \exp(a_0 + a_1 x)}$$

avec a_0 et a_1 des paramètres inconnus et à estimer. Ils expriment ensuite le taux de survie moyen en fonction de s comme une moyenne empirique. Ils intègrent ensuite cette expression dans la vraisemblance.

Un autre estimateur plus général a été proposé par Rogers-Bennett and Rogers (2006). Il exploite l'information continue apportée par la taille et n'a donc de sens que dans le cadre du modèle statistique continu. L'utilisation de l'information continue permet à cet estimateur d'être stable même pour des échantillons de petite taille. Cet estimateur est construit à partir d'une courbe de croissance $\Delta S = \varphi(S)$, où S est la taille et ΔS son accroissement sur un pas de temps. Il est défini en considérant la fonction de croissance ξ solution de l'équation différentielle :

$$\begin{cases} \dfrac{d\xi}{dt} = \varphi(\xi) \\ \xi(0) = u_0 \end{cases} \tag{1.14}$$

où u_0 est la taille minimale d'un individu de la population. La probabilité de passage dans la classe C_{i+1} estimée, $q_i^\bullet$ est alors une probabilité conditionnelle sachant que l'individu reste vivant. Soit u_i la borne supérieure de la classe C_i, δ_i sa largeur et ΔT la durée d'un pas de temps, elle s'exprime en fonction de ξ de la façon suivante :

$$q_i^\bullet = \frac{u_i - \xi[\xi^{-1}(u_i) - \Delta T]}{\delta_i}$$

La fonction φ est estimée par régression sur un échantillon d'individus conditionnellement au fait qu'ils restent vivants. On en déduit alors un estimateur $\hat{\xi}$ de ξ par l'équation 1.14. On obtient ainsi un estimateur de $q_i^\bullet$.

Un cas particulier de cet estimateur (Ralston et al., 2002 ; Namaalwa et al., 2005) est obtenu en considérant la moyenne empirique des accroissements en taille des individus de la classe C_i, notée $\overline{\Delta D}_i$, soit :

$$\hat{q}_i^\bullet = \frac{\overline{\Delta D}_i}{\delta_i} \tag{1.15}$$

Il correspond au cas où la fonction φ est estimée comme la moyenne empirique des accroissements par classe de taille :

$$\hat{\varphi}(x) = \sum_{i=1}^{I} \overline{\Delta D}_i \mathbf{1}_{u_{i-1} \leq x < u_i}$$

Ce type d'estimateur est particulièrement intéressant pour des petits échantillons car l'estimation est reliée au modèle statistique continu. Il permet de plus de définir les classes indépendamment de l'estimation et d'obtenir une estimation même quand des classes sont vides de données (ce que ne permet pas l'estimateur « classique »).

Dans cette thèse, nous nous focaliserons d'abord sur l'estimateur du maximum de vraisemblance du modèle discret, qui est l'estimateur le plus fréquemment employé dans les études forestières. Le « classicisme » de cet estimateur justifie qu'on s'y intéresse en premier, la suite immédiate du travail cherche à d'étendre les résultats obtenus pour cet estimateur à des estimateurs plus complexes (et plus robustes).

1.4.2 Estimateurs dans le modèle de Usher densité-dépendant

Dans le modèle forestier, n'importe quel taux de transition p, qu'il s'agisse d'un paramètre de transition ou d'une fécondité, est une fonction de la surface terrière, B, au temps courant, de l'effectif total au temps courant, N : $p = p_o\varphi(B, N, \theta)$ où θ est un paramètre inconnu. La fonction φ est une fonction donnée au départ telle que pour un état de référence de densité N_0 et de surface terrière B_0 : $\varphi(B_0, N_0, \theta) = 1$. La méthode classique pour estimer les taux de transition dans le modèle densité-dépendant consiste en une procédure en deux temps (Favrichon, 1998 ; Solomon, 1986). :

1. on estime les taux de transition pour chaque parcelle séparément ; soit $\hat{p}_k$ son estimateur pour la parcelle k
2. on ajuste le modèle $\hat{p}_k \sim p_o\varphi(B_k, N_k, \theta) + \epsilon_k$, où B_k est la surface terrière de la parcelle k et N_k sa densité.

Dans cette thèse, on s'intéresse à l'estimateur du maximum de vraisemblance de θ et on propose une estimation en une seule étape.

Chapitre 2

Distribution asymptotique des estimateurs des prédictions du modèle de Usher

Introduction

On s'intéresse, dans ce chapitre, au taux de croissance asymptotique λ_1 et à la distribution stationnaire w_1 définis dans le paragraphe 1.3.3 du chapitre 1. Le paramètre malthusien λ_1 est souvent utilisé pour savoir si la population est en voie d'extinction ($\lambda_1 < 1$) ou d'invasion ($\lambda_1 > 1$). Dans les applications forestières, le paramètre le plus significatif est la distribution stationnaire w_1, définie par la distribution diamétrique du peuplement. Une forêt naturelle non perturbée est, en effet, dans un système stationnaire, au moins pour une échelle de temps donnée.

L'apport de notre travail est de tenir compte de la variabilité d'échantillonnage. Dès lors que les taux de transition du modèle de Usher sont considérés comme des estimateurs, donc comme des variables aléatoires, les paramètres λ_1 et w_1 sont aussi des variables aléatoires. Le but de ce chapitre est alors de déterminer des intervalles de confiance autour des estimations de λ_1 et w_1. Houllier et al. (1989) ont déterminé une approximation de la variance et du biais d'un estimateur quelconque de λ_1, à l'aide d'un développement de Taylor. Dans ce chapitre, on établit un résultat asymptotique exact pour λ_1, et surtout pour w_1. On se focalise sur l'estimateur du maximum de vraisemblance et on en détermine la loi asymptotique. On obtient alors la variance asymptotique de λ_1 et w_1, et on montre qu'ils sont asymptotiquement non biaisés. Ces résultats fournissent directement un intervalle de confiance asymptotique pour λ_1 et un ellipsoïde de confiance asymptotique

pour w_1. On précise aussi la vitesse de convergence des estimateurs de λ_1 et w_1 vers cette loi à l'aide de simulations. Enfin, les résultats sont appliqués à un jeu de données récolté sur un site expérimental de Paracou en Guyane Française.

Les résultats présentés dans ce chapitre ont donné lieu à un article, accepté dans *Mathematical Biosciences*, dont on donne ici des compléments.

2.1 Loi asymptotique des estimateurs de λ_1 et w_1

Le paramètre λ_1 étant la première valeur propre de multiplicité un, et w_1 son vecteur propre associé, ils sont des fonctions des paramètres de U :

$$\lambda_1 = h(U) \quad \text{et} \quad w_1 = g\left(U, h(U)\right)$$

définies implicitement par :

$$\Phi(U, h(U)) = 0$$

où

$$\Phi(U, \lambda) = \det(U - \lambda I) \tag{2.1}$$

La fonction g est définie (modulo une constante multiplicative) par l'équation :

$$Ug(U, \lambda) - \lambda g(U, \lambda) = 0. \tag{2.2}$$

2.1.1 Estimateurs du maximum de vraisemblance

Les notations non définies ici sont reprises du chapitre 1 (et en particulier le paragraphe 1.1.1 du chapitre 1). Soit $d = (d_1, \ldots, d_I)$ la distribution en état de la population au temps initial t_0, avec $\sum_{i=1}^{I} d_i = 1$. Soit X un vecteur aléatoire qui prend ses valeurs dans l'ensemble $\mathcal{A} = \{(i, j, l); 1 \leq i \leq I-1; j = i, i+1; l \in \mathbf{N}\} \cup \{(I, I, l); l \in \mathbf{N}\} \cup \{(i, \dagger, l); 1 \leq i \leq I; l \in \mathbf{N}\}$, où $\dagger$ correspond à l'état des individus morts. Le vecteur aléatoire X décrit l'état des individus aux temps t_0 et $t_0 + 1$: (i, i, l) est un individu qui reste dans la classe i et donne naissance à l nouveaux individus ; $(i, i+1, l)$ est un individu qui passe dans la classe supérieure et donne naissance à l individus ; $(i, \dagger, l)$ est un individu qui donne naissance à l individus et qui meurt. La loi F de X est définie par :

$$\begin{aligned} \Pr[X = (i, i, l)] &= p_i \, d_i \, g_{il} \\ \Pr[X = (i, i+1, l)] &= q_{i+1} \, d_i \, g_{il} \\ \Pr[X = (i, \dagger, l)] &= (1 - p_i - q_{i+1}) d_i \, g_{il} \end{aligned}$$

où g_{il} est la probabilité qu'une variable aléatoire suivant une loi de Poisson de paramètre f_i soit égale à l :

$$g_{il} = \mathrm{e}^{-f_i} \frac{f_i^l}{l!}$$

Pour simplifier les notations, on note $\theta = (p_i, 1 \leq i \leq I; q_i, 2 \leq i \leq I; f_i, 1 \leq i \leq I; d_i, 1 \leq i \leq I) \in \mathbf{R}^{4I-1}$ le paramètre de la loi F.

Soit alors $(X_1, \ldots, X_n)$ un échantillon de taille n à valeurs dans $\mathcal{A}$, de loi $F(\theta)$. En pratique, cela correspond à considérer un échantillon de n individus observés à deux temps consécutifs, à partir duquel les paramètres de la matrice de Usher sont estimés. L'estimateur du maximum de vraisemblance de θ, $\widehat{\theta}_n$, correspond à une proportion et est donné, pour $i = 1 \ldots I$, par :

$$\widehat{p_i} = \frac{\sum_{k=1}^{n} \sum_{l=0}^{\infty} \mathbf{1}_{X_k=(i,i,l)}}{N_i}$$

$$\widehat{q_{i+1}} = \frac{\sum_{k=1}^{n} \sum_{l=0}^{\infty} \mathbf{1}_{X_k=(i,i+1,l)}}{N_i}$$

$$\widehat{f_i} = \frac{\sum_{k=1}^{n} \sum_{l=0}^{\infty} l \left(\mathbf{1}_{X_k=(i,i,l)} + \mathbf{1}_{X_k=(i,i+1,l)} + \mathbf{1}_{X_k=(i,\dagger,l)} \right)}{N_i}$$

$$\widehat{d_i} = \frac{N_i}{n}$$

où :

$$N_i = \sum_{k=1}^{n} \sum_{l=0}^{\infty} \left(\mathbf{1}_{X_k=(i,i,l)} + \mathbf{1}_{X_k=(i,i+1,l)} + \mathbf{1}_{X_k=(i,\dagger,l)} \right)$$

est le nombre d'individus présents dans la classe i au temps t_0. Dans le cas d'une estimation moyenne f, au lieu d'une fécondité par classe, l'estimateur du maximum de vraisemblance de f se simplifie en :

$$\hat{f} = \frac{\sum_{k=1}^{n} P(X_k)}{n}$$

où P est l'application de $\mathcal{A}$ dans $\mathbf{N}$ tel que $P[(i,j,l)] = l$. Le numérateur de $\hat{f}$ est, de façon basique, le nombre total de nouveaux individus engendrés par les n individus.

2.1.2 Comportement asymptotique des estimateurs du maximum de vraisemblance de λ_1 et w_1

La première valeur propre de U, λ_1, est une fonction implicite des paramètres de U : $\lambda_1 = h(U)$, où h est une application de $\mathbb{R}^{I\times I}$ dans $\mathbb{R}$. Pour simplifier, on identifie h à l'application de $\mathbb{R}^{3I-1}$ dans $\mathbb{R}$, telle que : $\lambda_1 = h(\Pi(\theta))$, Π étant la projection de $\mathbb{R}^{4I-1}$ dans $\mathbb{R}^{3I-1}$ défini par Π : $\theta = (p_i, q_i, f_i, d_i)_{1\leq i\leq I} \mapsto \Pi(\theta) = (p_i, q_i, f_i)_{1\leq i\leq I}$. De même, $w_1 = g(\Pi(\theta))$. Ainsi, l'estimateur du maximum de vraisemblance de λ_1 et de w_1 sont respectivement : $\hat{\lambda_1}_n = h(\Pi(\hat{\theta}_n))$ et $\hat{w_1}_n = g(\Pi(\hat{\theta}_n))$. Le comportement asymptotique de $\hat{\lambda_1}_n$ et $\hat{w_1}_n$ sont donnés par la convergence en loi suivante :

Proposition 1 *Avec les notations ci-dessus,*

$$\begin{aligned}\sqrt{n}\left(\hat{\lambda_1}_n - \lambda_1\right) &\xrightarrow{\mathcal{L}} \mathcal{N}\big(0,\ (d_{\Pi(\theta)}h)^t.(d_\theta\Pi)^t.J(\theta)^{-1}.(d_\theta\Pi).(d_{\Pi(\theta)}h)\big)\\ \sqrt{n}\left(\hat{w_1}_n - w_1\right) &\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (d_{\Pi(\theta)}g)^t.(d_\theta\Pi^t).J(\theta)^{-1}.(d_\theta\Pi).(d_{\Pi(\theta)}g)\right)\end{aligned}$$

où $\xrightarrow{\mathcal{L}}$ correspond à la convergence en loi, M^t est la transposée de la matrice M, $d_a h$ est la différentielle de h en a, et le point indique la multiplication matricielle. De plus, $J(\theta)$ est l'information de Fisher dont l'inverse est égale à :

$$\begin{aligned}
&(J(\theta)^{-1})_{i,i} = \frac{p_i(1-p_i)}{d_i} && \text{pour } i = 1,\dots,I\\
&(J(\theta)^{-1})_{I+i,i} = (J(\theta)^{-1})_{i,I+i} = -\frac{p_i q_{i+1}}{d_i} && \text{pour } i = 1,\dots,I-1\\
&(J(\theta)^{-1})_{I+i,I+i} = \frac{q_{i+1}(1-q_{i+1})}{d_i} && \text{pour } i = 1,\dots,I-1\\
&(J(\theta)^{-1})_{2I-1+i,2I-1+i} = \frac{f_i}{d_i} && \text{pour } i = 1,\dots,I\\
&(J(\theta)^{-1})_{3I-1+i,3I-1+i} = d_i(1-d_i) && \text{pour } i = 1,\dots,I\\
&(J(\theta)^{-1})_{3I-1+i,3I-1+j} = (J(\theta)^{-1})_{3I-1+j,3I-1+i} = -d_i\,d_j && \text{pour } i = 1,\dots,I; j = 1,\dots,I; i\neq j
\end{aligned}$$

La preuve est donnée dans l'annexe A. De plus, les différentielles de h et g sont calculées dans l'annexe du chapitre.

2.2 Simulations et application

2.2.1 Application

Les résultats sur le comportement des estimateurs de λ_1 et w_1 sont appliqués à un jeu de données récolté sur un site expérimental forestier à Paracou, en Guyane Française. Le site, ainsi que les données, sont décrits dans l'annexe A. En particulier, les données sont prises sur un pas de temps égal à deux ans, à des dates où le peuplement forestier n'a subi aucune exploitation. De plus, tous les arbres du site ont été répertoriés et sont au nombre de 47 301.

A partir de ces données, on obtient des estimations des paramètres de la matrice de Usher, ainsi que de λ_1 et w_1. En utilisant la proposition 1, on a une estimation de la variance asymptotique de λ_1 et de la matrice de covariance de w_1, en remplaçant θ par $\hat{\theta}_n$ dans l'expression de la variance asymptotique. On obtient directement un intervalle de confiance asymptotique pour λ_1, et une ellipsoïde de confiance pour w_1 dans $\mathbb{R}^{I-1}$, au niveau 5%.

Les résultats montrent en particulier que λ_1 est légèrement supérieur à 1. De plus, la valeur 1 n'appartient pas à l'intervalle de confiance. Il est pourtant peu probable que le peuplement forestier va croître de façon exponentielle. Ce résultat peut provenir, d'une part, de conditions environnementales favorables passagères ou d'une exploitation récente, et, d'autre part, des simplifications du modèle (en particulier, l'hypothèse de stationnarité).

L'interprétation de l'ellipsoïde de confiance de w_1 est moins directe. On peut en déduire immédiatement des intervalles de confiance par classe, mais ils ne prennent pas en compte les interactions entre les classes. Une discussion sur l'interprétation de l'ellipsoïde de confiance de w_1 est donnée dans l'annexe A.

Enfin, un autre résultat important est que les comportements asymptotiques de λ_1 et w_1 fournissent le nombre d'arbres à inventorier pour obtenir une précision donnée sur l'estimation. Par exemple, pour avoir une précision d'au moins 10% sur l'estimation de w_1, la taille de l'échantillon doit être supérieure à 8006 arbres.

2.2.2 Simulations

Les lois des estimateurs de λ_1 et w_1 sont obtenues de façon asymptotique. Pour préciser la vitesse de convergence des estimateurs de λ_1 et w_1, on utilise des simulations. Elles permettent de savoir à partir de quelles tailles d'échantillon les résultats asymptotiques s'appliquent. On se focalise sur deux quantités :

– la variance de λ_1, notée γ ;

– la racine carrée de la première valeur propre de la matrice de covariance de w_1, notée μ.

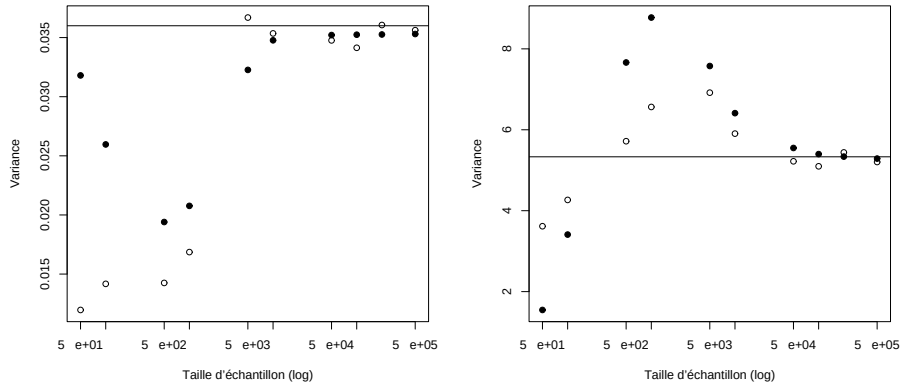
Cette dernière est utilisée car elle est proportionnelle à la longueur du plus grand axe principal de l'ellipsoïde de confiance. Les paramètres de simulation sont les paramètres de la loi F obtenus à Paracou. Pour une taille d'échantillon n donnée, on simule 1000 échantillons de taille n de loi F. Pour chaque taille d'échantillon n, une estimation $\hat{\theta}_n$ est calculée. Les valeurs asymptotiques théoriques sont comparées aux valeurs simulées pour savoir, d'une part, pour quelles tailles d'échantillon la variance asymptotique de λ_1 et la matrice de covariance de w_1 sont bien estimées, et d'autre part, pour quelles tailles d'échantillon le résultat asymptotique s'applique. Pour chaque taille n d'échantillon :

1. on obtient 1000 estimations de γ et de μ. On trace ensuite, par un point, la moyenne empirique de ces estimations ;
2. on obtient 1000 estimations de λ_1 et de w_1. On calcule ensuite la variance empirique de λ_1, $\hat{s_n}$ et la matrice de covariance empirique de w_1, $\hat{S}_n$ à partir de ces valeurs. On note $\hat{\mu}_n$ la racine carrée de la première valeur propre de $\hat{S}_n$. On trace ensuite les fonctions $n \mapsto n\,\hat{s}_n$ et $n \mapsto \sqrt{n}\,\hat{\mu}_n$.

Dans les simulations, la taille de l'échantillon varie de 50 à 500 000.

Les résultats des simulations sont données par les figures 2 et 3 de l'annexe A et y sont commentés. Ils confirment en particulier les comportements asymptotiques des estimateurs et mettent en évidence que ces résultats théoriques s'appliquent pour de grandes tailles d'échantillon (supérieures à 50000). Par contre, ils sont peu adaptés à l'étude d'espèces peu abondantes.

Par ailleurs, dans les applications, la vraie valeur de la variance des prédictions n'étant pas accessible, l'estimation de la variance asymptotique est la valeur que l'on considère. On compare alors les valeurs obtenues dans les simulations pour la variance empirique avec celles pour la variance asymptotique estimée. Les résultats sont donnés dans la figure 2.1. La figure montre tout d'abord que la valeur de la variance asymptotique estimée est proche de la valeur de la variance empirique pour des grandes tailles d'échantillon : supérieures à 10 000 pour λ_1 et supérieures à 50 000 pour w_1. Ces résultats confirment les conclusions des simulations précédentes. Toutefois, pour des tailles d'échantillon « moyennes », l'écart entre les valeurs pour la variance asymptotique estimée et la variance empirique est peu importante : cet écart est de l'ordre de 10% de la valeur de la vraie variance asymptotique. Ceci est constaté pour des tailles d'échantillon comprises entre 500 et 10 000 pour λ_1 et comprises entre 5 000 et 50 000 pour w_1.



(a) (b)

FIG. 2.1 – Comparaison de la variance empirique et de l'estimation de la variance asymptotique pour (a) λ_1, et (b) w_1 (1000 simulations). Les points noirs correspondent à la variance asymptotique estimée alors que les points blancs correspondent à la variance empirique. La droite horizontale correspond à la valeur de la vraie variance asymptotique

Par ailleurs, la variance asymptotique estimée de λ_1 est plus grande que sa variance empirique pour des tailles d'échantillon inférieures à 1 000, puis est plus petite pour des tailles d'échantillon supérieures jusqu'à ce que l'écart entre ces deux valeurs se stabilise. Pour w_1, la valeur correspondant à la matrice de covariance asymptotique est plus petite que celle correspondant à la matrice de covariance empirique pour des tailles d'échantillon inférieures à 500, puis est plus grande jusqu'à ce que l'écart entre ces deux valeurs se stabilise. L'estimation de la variance asymptotique sur-estime donc la variance de λ_1 pour des petites tailles d'échantillon et la sous-estime pour des tailles d'échantillon supérieures. A contrario, l'estimation de la matrice de covariance asymptotique sous-estime la matrice de covariance de w_1 pour des petites tailles d'échantillon et la sur-estime pour des tailles d'échantillon supérieures.

Ces résultats nuancent donc les résultats des précédentes simulations pour des tailles d'échantillon « moyennes », et précisent la façon dont on estime la variance asymptotique par rapport à la variance des estimateurs de λ_1 et w_1.

2.2.3 Comparaison de la variabilité démographique et d'échantillonnage

Dans cette partie, la variabilité des prédictions du modèle de Usher a été étudiée en tenant compte de la variabilité due à l'échantillonnage. La stochasticité démographique n'a pas été considérée ici. Pour comparer le poids de la stochasticité d'échantillonnage par rapport à la stochasticité démographique sur la variabilité des prédictions, on simule 1000 chaînes de Markov sur $\mathbb{N}^I$ décrivant l'évolution du vecteur d'effectif, de loi initiale le vecteur de distribution en 1984 et de fonction de transition déterminée par les taux de transition obtenus à Paracou. Pour une taille initiale de la population n, on simule les trajectoires des chaînes de Markov jusqu'à un temps T. On obtient alors 1000 vecteurs d'effectifs au temps T. On calcule la variance empirique de ces 1000 vecteurs d'effectifs divisé par leur norme. On la compare avec la matrice de covariance asymptotique de w_1 à Paracou.

Pour $n = 46470$, correspondant à la taille de la population à Paracou en 1984, et pour $T = 100$, on obtient la matrice de covariance empirique du vecteur d'effectif normalisé, au temps T, multipliée par n, donnée dans la table 2.1.

La matrice de covariance obtenue par 1000 simulations de chaînes de Markov est 100 fois plus petite que la matrice de covariance asymptotique de w_1 à Paracou. On met ainsi en évidence le poids de la stochasticité d'échantillonnage dans la variabilité des prédictions du modèle de Usher.

2.3 Conclusion

Dans ce chapitre, nous nous sommes focalisés sur l'estimateur du maximum de vraisemblance. Le comportement asymptotique de cet estimateur a permis, à l'aide de la delta-méthode, de construire des intervalles de confiance asymptotiques des prédictions. Nous aspirons alors à avoir des intervalles de confiance optimaux, c'est-à-dire dont la longueur est la plus petite possible et en terme de probabilité de couverture la plus grande possible. Affiner les intervalles de confiance va être l'objet du chapitre suivant qui se focalise sur la recherche d'estimateurs des taux de transition robustes.

Par ailleurs, les résultats appliqués aux données de Paracou ont abouti à des conclusions peu réalistes. La valeur propre λ_1 était, en effet, légèrement plus grande que 1, pour un niveau de confiance égal à 5%. Ceci signifie que la population explose à long terme. En fait, ces résultats peuvent provenir du fait que les conditions environnementales étaient favorables pendant le pas de temps considéré. Le modèle de Usher linéaire fait en effet l'hypothèse que

les paramètres sont indépendants du temps. Cette hypothèse est relâchée par l'utilisation des modèles densité-dépendants. Le chapitre 4 sera alors consacré à étendre les résultats de ce chapitre aux modèles densité-dépendants.

TAB. 2.1 – Matrice de covariance, multipliée par $n = 46470$, du vecteur d'effectif normalisé, noté $\widehat{N}$, obtenu par 1000 simulations d'une chaîne de Markov de taille initiale n, au pas de temps $T = 100$. $\widehat{N}_i$ $(i = 1 \dots m)$est la $i$ème composante de $\widehat{N}$.

	$\widehat{N}_1$	$\widehat{N}_2$	$\widehat{N}_3$	$\widehat{N}_4$	$\widehat{N}_5$
$\widehat{N}_1$	0.1858				
$\widehat{N}_2$	$-6.1637 \cdot 10^{-2}$	0.1450			
$\widehat{N}_3$	$-3.4570 \cdot 10^{-2}$	$-2.6365 \cdot 10^{-2}$	$9.1411 \cdot 10^{-2}$		
$\widehat{N}_4$	$-2.1515 \cdot 10^{-2}$	$-1.5524 \cdot 10^{-2}$	$-1.2156 \cdot 10^{-2}$	$6.3785 \cdot 10^{-2}$	
$\widehat{N}_5$	$-2.0030 \cdot 10^{-2}$	$-9.955 \cdot 10^{-3}$	$-2.2828 \cdot 10^{-3}$	$-8.5740 \cdot 10^{-3}$	$4.8829 \cdot 10^{-2}$
$\widehat{N}_6$	$-1.0875 \cdot 10^{-2}$	$-1.3399 \cdot 10^{-2}$	$-4.9943 \cdot 10^{-3}$	$-1.5630 \cdot 10^{-3}$	$-2.5631 \cdot 10^{-3}$
$\widehat{N}_7$	$-7.1620 \cdot 10^{-3}$	$-5.1765 \cdot 10^{-3}$	$-3.0666 \cdot 10^{-3}$	$-1.5798 \cdot 10^{-3}$	$-1.1470 \cdot 10^{-3}$
$\widehat{N}_8$	$-5.4644 \cdot 10^{-3}$	$-5.1310 \cdot 10^{-3}$	$-9.4825 \cdot 10^{-4}$	$-1.2457 \cdot 10^{-3}$	$-8.2481 \cdot 10^{-4}$
$\widehat{N}_9$	$-6.7622 \cdot 10^{-3}$	$-2.0109 \cdot 10^{-3}$	$-1.2857 \cdot 10^{-3}$	$-4.1320 \cdot 10^{-4}$	$-1.4094 \cdot 10^{-3}$
$\widehat{N}_{10}$	$-6.2783 \cdot 10^{-3}$	$8.1131 \cdot 10^{-4}$	$-2.2439 \cdot 10^{-3}$	$-2.5366 \cdot 10^{-4}$	$-2.0135 \cdot 10^{-3}$
$\widehat{N}_{11}$	$-1.1492 \cdot 10^{-2}$	$-6.6223 \cdot 10^{-3}$	$-3.4978 \cdot 10^{-3}$	$-9.5934 \cdot 10^{-4}$	$-2.9371 \cdot 10^{-5}$

	$\widehat{N}_6$	$\widehat{N}_7$	$\widehat{N}_8$	$\widehat{N}_9$	$\widehat{N}_{10}$	
$\widehat{N}_1$						
$\widehat{N}_2$						
$\widehat{N}_3$						
$\widehat{N}_4$						
$\widehat{N}_5$						
$\widehat{N}_6$	$3.8411 \cdot 10^{-2}$					
$\widehat{N}_7$	$-1.4035 \cdot 10^{-3}$	$2.2025 \cdot 10^{-2}$				
$\widehat{N}_8$	$-2.2479 \cdot 10^{-3}$	$-9.5599 \cdot 10^{-4}$	$1.9081 \cdot 10^{-2}$			
$\widehat{N}_9$	$-1.2139 \cdot 10^{-4}$	$-2.4837 \cdot 10^{-4}$	$-1.1388 \cdot 10^{-3}$	$1.3630 \cdot 10^{-2}$		
$\widehat{N}_{10}$	$1.6365 \cdot 10^{-4}$	$-3.6844 \cdot 10^{-4}$	$1.8188 \cdot 10^{-5}$	$-3.5631 \cdot 10^{-5}$	$1.0551 \cdot 10^{-2}$	
$\widehat{N}_{11}$	$-1.4079 \cdot 10^{-3}$	$-9.1681 \cdot 10^{-4}$	$-1.1419 \cdot 10^{-3}$	$-2.0406 \cdot 10^{-4}$	$-3.5094 \cdot 10^{-4}$	2.6621

Chapitre 3

Robustesse des taux de transition

Introduction

La robustesse se place dans le cadre du choix d'un modèle statistique (Huber, 2004 ; SociétéMathématiqueDeFrance, 1977). Le modèle étant une simplification de phénomènes réels, des erreurs de mesures peuvent intervenir. Elles peuvent être, par exemple, des erreurs d'expérimentation, dues à la sensibilité du matériel, des erreurs grossières (mesures aberrantes, etc.). Ces erreurs peuvent entraîner dans les conclusions finales des écarts par rapport à la réalité. Une procédure robuste consiste à rendre insensible les estimateurs des paramètres d'intérêt à ces petites déviations en perturbant le modèle d'échantillonnage. On peut, par exemple, modéliser les erreurs grossières par une loi de grande variance. Dans le cas d'un modèle gaussien $\mathcal{N}(\theta, \sigma^2)$, où on veut estimer la moyenne θ, avec σ^2 connu, on peut, pour tenir compte des mesures aberrantes, choisir le modèle :

$$(1-\epsilon)\mathcal{N}(\theta, \sigma^2) + \epsilon\mathcal{N}(\theta, k\sigma^2) \qquad \text{avec } k > 1,\ \epsilon > 0 \text{ constantes fixées}$$

Ce modèle a la même moyenne que le modèle initial, mais la variance est perturbée par un facteur k.

La recherche de procédure robuste est guidée par les contraintes du modèle. En particulier, la structure discrète et l'hypothèse de Usher imposent des contraintes sur la dynamique de la population : l'évolution des individus est décrite suivant leur classe d'état, et les transitions entre les classes sont limitées. Ces contraintes influent sur les prédictions du modèle. On se focalise alors sur deux questions :

– comment rendre l'hypothèse de Usher (définie au chapitre 1) réaliste, autrement dit faire en sorte qu'il y ait peu de données qui violent cette hypothèse ?

– l'hypothèse de Usher étant vérifiée par les données, les erreurs de mesures entraînent des erreurs sur la classification des individus dans leur classe d'état. Comment minimiser l'impact de ces erreurs sur la dynamique de la population ?

La validité de l'hypothèse de Usher et les erreurs de classification des individus dépendent du choix des classes d'état, la validité de l'hypothèse de Usher dépendant aussi du pas de temps choisi. Ces deux questions peuvent alors être traitées en déterminant les classes d'état et le pas de temps adéquats, et cela en fonction des données. Cette approche sera discutée dans le chapitre de conclusion. On suppose ici que ces paramètres sont fixés et on adopte un autre point de vue qui est la robustesse des estimateurs des taux de transition.

On propose deux façons de traiter les données qui ne respectent pas l'hypothèse de Usher :

– les éliminer, solution fréquemment retenue par les praticiens ;
– les lisser de la façon suivante : les individus qui passent de plus d'une classe pendant le pas de temps sont considérés comme des individus qui passent dans la classe supérieure, et les individus qui régressent de classe sont considérés comme des individus qui restent dans leur classe.

Une autre solution pour autoriser plus de transitions entre les classes d'état a été proposée en utilisant le modèle de Lefkovitch par exemple (Solomon, 1986 ; Lamar and McGraw, 2005), mais elle ne s'avère pas satisfaisante lorsque le nombre de données est insuffisant. Des deux façons de traiter les données ici on obtient deux estimateurs des taux de transition, suivant le traitement des données choisi : un estimateur tronqué, dans le cas où on élimine les données, et un estimateur non tronqué dans le cas où on les lisse. Dans ce chapitre, on comparera alors ces deux estimateurs.

Les erreurs de classification des individus proviennent des erreurs de mesure sur la taille et l'accroissement en taille des individus. On envisage alors deux façons de minimiser ces erreurs, dans le sens où elles ont un impact sur la dynamique de la population :

– on élimine les individus qui ont les plus grands accroissements en taille pendant un pas de temps. En effet, les grands accroissements sont considérés comme des données aberrantes, donc comme des erreurs de mesure ;
– on élimine les individus dont les états à un temps donné et au temps suivant appartiennent aux bords des classes.

On compare alors trois estimateurs : l'estimateur classique, où on garde les données qui respectent l'hypothèse de Usher, l'estimateur tronqué provenant de la troncature sur les accroissements et celui provenant de la troncature aux bornes des classes.

Pour chacune des deux questions, on cherche l'estimateur le plus robuste, parmi les estimateurs ainsi construits. Pour quantifier et comparer la robustesse des estimateurs, on utilise la notion de fonction d'influence. Celle-ci donne en effet des critères quantitatifs de robustesse, qui sont basées sur des notions de sensibilité des estimateurs. La fonction d'influence permet aussi de déterminer le comportement asymptotique des estimateurs. Bien que les fonctionnelles des estimateurs considérés s'éexpriment de manière bivariée (l'une des variables étant la taille des individus et l'autre leur accroissement en taille) nous nous ramènerons à étudier des fonctionnelles univariées de quantiles pour éviter des questions de définition de quantiles dans le cas bivarié (Wang and Serfling, 2006). On s'inspirera alors de la théorie sur les L-estimateurs liée à la fonction d'influence pour construire les estimateurs proposés et quantifier leur robustesse.

Ce chapitre est structuré en deux parties. La première partie présente des résultats généraux sur les L-estimateurs, qui seront nécessaires pour la seconde partie. La seconde partie se focalise sur les estimateurs des taux de transition du modèle d'Usher, qui sont des cas particuliers de L-estimateurs. On commencera par des rappels sur les outils nécessaires à l'étude de la robustesse (fonction d'influence, sensibilité, L-estimateur...), puis on donnera la démonstration de deux lemmes permettant le calcul de la fonction d'influence des estimateurs qui nous intéressent. La seconde partie débutera par la définition des estimateurs des taux de transition du modèle de Usher, puis on étudiera la robustesse de ces différents estimateurs.

Ce chapitre est un complément d'article, accepté dans *Computational Statistics and Data Analysis* et présenté en annexe C.

Notations

Si F est une fonction de répartition, F n'étant pas continue en général, on définit sa fonction réciproque F^{-1} par la formule :

$$F^{-1}(t) = \inf\{x; F(x) \geq t\}$$

pour $t \in]0, 1[$.

3.1 Etude générale de la robustesse

On s'intéresse, dans ce travail, au cas particulier de l'estimation d'un paramètre de localisation. La théorie de la robustesse dans ce cas utilisant la notion de fonction d'influence a été développé principalement par Hampel (1974, 1986) et Huber (1964, 2004).

Dans cette partie, on considère un n-échantillon $(X_1, \ldots, X_n)$ à valeurs dans $(E, \mathcal{E})$, un espace borélien. On note $F(\theta)$ la loi de cet échantillon, où θ est un paramètre inconnu, et F_n sa fonction de répartition empirique. On suppose de plus que $(E, \|.\|_E)$ est un espace normé.

3.1.1 Estimateurs robustes asymptotiquement efficaces

Dans le cas où F n'est qu'approximativement connue, on peut considérer que F appartient à un voisinage $\mathcal{L}$ d'une loi G connue. $\mathcal{L}$ peut être par exemple :

$$\mathcal{L} = \{(1-\epsilon)G + \epsilon H, H \in \mathcal{M}\}$$

où $\mathcal{M}$ est une famille de lois et $0 < \epsilon < 1$. G correspond à la fonction de répartition du modèle initialement choisi.

Dans le modèle initial, l'estimateur du maximum de vraisemblance est asymptotiquement efficace. Mais il est peu robuste, en ce sens que sa variance asymptotique n'est pas bornée, en général, dans le domaine $\mathcal{M}$.

Trouver des estimateurs robustes consiste alors à trouver des estimateurs qui donnent de bons résultats dans le modèle initial au sens de l'efficacité, et qui restent peu sensibles à de légers écarts par rapport à ce modèle (Huber, 2004 ; SociétéMathématiqueDeFrance, 1977).

Dans le modèle de Usher, où la loi des paramètres est discrète, l'estimateur du maximum de vraisemblance correspond à l'estimateur empirique de la moyenne. Pour robustifier cet estimateur on peut considérer la médiane. Ces deux estimateurs font partie de la classe des M-estimateurs, définis, dans le cas d'un estimateur de localisation, par l'ensemble des statistiques $T_n(X_1, \ldots, X_n)$ solution de :

$$\Sigma_{i=1}^{n} \Psi\left(X_i - T\right) = 0$$

où Ψ est l'estimateur de Huber, qui correspond aux fonctions :

$$\Psi_k : t \mapsto \begin{cases} t & \mid t \mid \leq k \\ k\text{sign}(t) & \mid t \mid > k \end{cases}$$

Cet estimateur fait le pont entre la moyenne empirique, qui est le cas limite $k \to \infty$, et la médiane, qui est le cas limite $k \to 0$. Les estimateurs considéré dans ce chapitre étant des estimateurs construits à partir de données tronquées, on s'intéressera plutôt ici à la classe des L-estimateurs, dont un cas particulier est l'estimateur α-tronqué. Ce dernier consiste à éliminer dans un

n-échantillon ordonné les $[n\alpha]$ plus petites et les $[n\alpha]$ plus grandes observations, où $[x]$ désigne la partie entière de x. Ces estimateurs sont développés au paragraphe 3.1.3.

D'autre part, la robustesse des estimateurs peut être quantifiée de façon numérique par la notion de fonction d'influence. En particulier, la fonction d'influence, dont la définition est rappelée dans le paragraphe suivant, donne des critères de robustesse d'estimateurs.

3.1.2 Fonction d'influence

Soit $\mathcal{F}$ l'espace des fonctions de distribution et $\mathcal{F}_n$ l'ensemble des probabilités concentrées sur un ensemble de n points de E et uniformes : $\mathcal{F}_n = \left\{\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}; (x_i)_{1\leq i\leq n} \in E^n\right\}$. δ_x dénote la distribution de Dirac en x. Soit $T_n = T_n(X_1, \ldots, X_n)$ un estimateur.

Soit T une application de $\mathcal{F}_0 \subset \mathcal{F}$ dans E telle que $\mathcal{F}_0$ contient F et $\mathcal{F}_n$. On suppose que la séquence T_n est générée par la fonctionnelle T, c'est-à-dire que :

$$T_n = T_n(X_1, \ldots, X_n) = T(F_n)$$

Plusieurs estimateurs vérifient cette condition. C'est le cas, par exemple, des M-estimateurs (équation (3.1)) et des L-estimateurs (équation (3.2)) (Huber, 2004 ; Boos and Serfling, 1980 ; Hampel, 1974), dans le cas où $E = \mathbb{R}$, dont les fonctionnelles $T(F)$ vérifient :

$$\int \psi(s - T(F))F(ds) = 0 \tag{3.1}$$

$$T(F) = \int h(F^{-1}(s))m(s)ds \tag{3.2}$$

où ψ et h sont des fonctions réelles et m est une densité.

Différentiabilité au sens de Fréchet et de Gâteaux

L'utilisation des propriétés de différentiabilité des fonctionnelles T permet d'obtenir des résultats statistiques pour la séquence (T_n). L'idée d'utiliser ces propriétés a été initiée par Von Mises (1947) ; Kallianpur and Rao (1955), mais a réellement pris son essor par le développement des théories sur la robustesse. Une illustration importante est la fonction d'influence (Hampel, 1974), définie comme une différentielle de T. De plus, pour de bonnes propriétés de différentiabilité de T, elle détermine la loi asymptotique de T (Boos and Serfling, 1980 ; Huber, 2004).

Soit $\mathcal{F}$ l'espace des fonctions de distribution et $\mathcal{D}$ l'espace affine généré par les différences $G - H$ d'éléments de $\mathcal{F}$. On munit $\mathcal{D}$ de la métrique $\|.\|$.

On dit qu'une fonctionnelle T définie sur $\mathcal{F}$ est *Fréchet-différentiable* au point F de $\mathcal{F}$ s'il existe une fonctionnelle linéaire continue, L_F, définie sur $\mathcal{F}$ telle que pour tout G dans $\mathcal{F}$:

$$\|T(G) - T(F) - L_F(G - F)\|_E = o\left(\|G - F\|\right)$$

Si, de plus, T est faiblement continue, L_F est linéaire faiblement continue, et telle que :

$$L_F(G - F) = \int \psi_F dG \tag{3.3}$$

où ψ_F est une fonction bornée et continue, telle que : $\int \psi_F dF = 0$. C'est le cas par exemple des M- et L-estimateurs.

Une notion plus faible de différentiabilité est la différentiabilité au sens de Gâteaux. C'est une différentiabilité au sens « directionnel ». Une fonctionnelle T définie sur $\mathcal{F}$ est dite *Gâteaux-différentiable* au point F de $\mathcal{F}$ s'il existe une fonctionnelle linéaire continue, L_F, définie sur $\mathcal{F}$ telle que pour tout G dans $\mathcal{F}$:

$$T[F + \epsilon(G - F)] - T(F) - \epsilon\, L_F(G - F) = o\left(\epsilon\right) \tag{3.4}$$

La propriété de différentiabilité au sens de Gâteux permet de définir la fonction d'influence.

Une notion intermédiaire de différentiabilité qui semble plus adaptée dans ce chapitre est la notion d'*Hadamard-différentiabilité* qui consiste à étudier la Gâteaux-différentiabilité uniformément sur les compacts.

Définition

Sous certaines conditions de régularité que l'on va préciser, l'estimateur T_n a des propriétés de consistance et d'efficacité. La consistance de l'estimateur T_n a été établie par Hampel (1971). Si T est faiblement continue en F, alors la suite $(T_n)_n$ est consistante en F, au sens où $T_n \overset{n\to\infty}{\to} T(F)$ en probabilité et presque-sûrement.

On suppose que T est Gâteaux-différentiable. Étant donné x dans E, on pose $G = \delta_x$ dans l'équation (3.4). La fonction d'influence correspond à la différentielle L_F dans (3.4). Plus précisément, on obtient la définition suivante (Hampel, 1974, 1986) :

Définition 1 *La fonction d'influence au point x, $IC_{T,F}(x)$, de T en F suivant la direction δ_x est égale à :*

$$IC_{T,F}(x) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon}$$

Si, de plus, la suite (T_n) est consistante, alors la quantité $\epsilon IC_{T,F}(x)$ fournit, pour ϵ assez petit, une valeur approchée du biais asymptotique de l'estimateur T_n, causé par une contamination de F par une masse ϵ.

Par ailleurs, si T est différentiable au sens de Fréchet ou d'Hadamard au point F (Boos and Serfling, 1980 ; Rieder, 1994), alors la fonction d'influence correspond à la fonction ψ_F dans l'équation (3.3) et le comportement asymptotique de (T_n) est donné par :

$$\sqrt{n}(T(F_n) - T(F)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \int IC_{T,F}(x)^2 dx\right)$$

où $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

Robustesse dans le cadre de la fonction d'influence

La fonction d'influence permet d'obtenir des critères quantitatifs pour mesurer la robustesse d'un estimateur.

Un premier critère est donné par la borne supérieure de la norme de la fonction d'influence. On définit ainsi la sensibilité globale :

$$\mu = \sup_x \|IC_{T,F}(x)\|_\infty$$

où $\|.\|_\infty$ est la norme infinie. Elle mesure la plus grosse influence causée par une petite contamination sur la valeur de l'estimateur.

Un deuxième critère concerne les petites fluctuations sur les observations. On définit la sensibilité locale par :

$$\nu = \sup_{x \neq y} \frac{\|IC_{T,F}(y) - IC_{T,F}(x)\|_\infty}{\|y - x\|_\infty}$$

Elle mesure l'effet sur la statistique lorsqu'une observation x est remplacée par une autre y. En particulier, elle permet d'évaluer les effets de regroupements sur la statistique.

Dans ce chapitre, on s'intéresse à des estimateurs tronqués dans le cas multidimensionnel. On rappelle dans le paragraphe suivant l'expression de la fonction d'influence de la moyenne α-tronquée et plus généralement des L-estimateurs dans le cas unidimensionnel. L'extension de ces résultats au cas multidimensionnel reposera sur deux lemmes qui seront démontrés dans le paragraphe 3.1.4.

3.1.3 L-estimateurs : cas unidimensionnel

Dans ce paragraphe, on se place dans le cas où $E = \mathbb{R}$.

Cas général des L-estimateurs

Soit $(X_{(1)}, \ldots, X_{(n)})$ l'échantillon ordonné, avec $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. On considère une statistique comme étant une combinaison linéaire des $X_{(i)}$ ou plus généralement d'une fonction h des $X_{(i)}$:

$$T_n = \sum_{k=1}^{n} a_{n,k} h(X_{(k)})$$

On suppose que les $a_{n,k}$ sont engendrées par une mesure M sur $[0;1]$, de densité m.

Alors, $T_n = T(F_n)$ est engendrée par la fonctionnelle (Huber, 2004) :

$$T(F) = \int_0^1 h\left(F^{-1}(s)\right) m(s) \mathrm{d}s \tag{3.5}$$

La fonction d'influence, $IC_{T,F}(x)$, de T au point F dans la direction x est la dérivée en 0 de la fonction $t \mapsto T(F_t)$, où $F_t = (1-t)F + t\delta_x$. Soit S le support de m et V un voisinage de 0. On suppose que h est dérivable sur un intervalle contenant $F_t^{-1}(S)$, pour tout $t \in V$. Alors, $IC_{T,F}(x)$ s'écrit (Huber, 2004) :

$$IC_{T,F}(x) = \int_{-\infty}^{x} h'(y) m(F(y)) \mathrm{d}y - \int_{-\infty}^{+\infty} [1 - F(y)]\, h'(y) m(F(y)) \mathrm{d}y \tag{3.6}$$

La condition sur h permet de légitimer la différentiation de la fonction $t \mapsto h(F_t^{-1}(s))$ sous le signe intégral.

Cas particulier : moyenne α-tronquée

Soit $0 < \alpha < 1/2$. La moyenne α-tronquée correspond, dans l'équation (3.5), à $h(x) = x$ et :

$$\begin{aligned} m(s) &= \frac{1}{1-2\alpha}, && \text{pour } \alpha < s < 1-\alpha \\ &= 0, && \text{sinon.} \end{aligned}$$

Ainsi :

$$T(F) = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} F^{-1}(s) \mathrm{d}s$$

L'estimateur s'écrit alors :

$$T(F_n) = \frac{1}{n(1-2\alpha)} \left[\sum_{k=[n\alpha]+2}^{n-[n\alpha]-1} X_{(k)} + (1 - n\alpha + [n\alpha]) \left(X_{([n\alpha]+1)} + X_{(n-[n\alpha])} \right) \right]$$

où $[x]$ indique la partie entière de x. On remarque que si $n\alpha$ n'est pas un entier, c'est-à-dire $n\alpha = [n\alpha] + p$, les observations $X_{([n\alpha]+1)}$ et $X_{(n-[n\alpha])}$ sont affectées d'un poids égal à $1-p$.

D'après l'équation (3.6), la fonction d'influence de la moyenne α-tronquée est :

$$\begin{aligned} IC_{T,F}(x) &= \frac{1}{1-2\alpha}\left[F^{-1}(\alpha) - W(F)\right] && \text{pour } x < F^{-1}(\alpha) \\ &= \frac{1}{1-2\alpha}\left[x - W(F)\right] && \text{pour } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ &= \frac{1}{1-2\alpha}\left[F^{-1}(1-\alpha) - W(F)\right] && \text{pour } x > F^{-1}(1-\alpha) \end{aligned}$$

avec :

$$\begin{aligned} W(F) &= \int_{\alpha}^{1-\alpha} F^{-1}(s)\mathrm{d}s + \alpha F^{-1}(\alpha) + \alpha F^{-1}(1-\alpha) \\ &= (1-2\alpha)T(F) + \alpha F^{-1}(\alpha) + \alpha F^{-1}(1-\alpha) \end{aligned}$$

W est appelée la moyenne α-Winzorisée.

Remarque 1 *: On peut généraliser la moyenne α-tronquée de la façon suivante : soit $0 < a < 1$, et soit $0 < \alpha_1 < \dfrac{a}{2}$ et $0 < \alpha_2 < \dfrac{1-a}{2}$. On considère l'estimateur engendré par :*

$$T(F) = \frac{1}{1-2(\alpha_1+\alpha_2)}\left[\int_{\alpha_1}^{a-\alpha_1} F^{-1}(s)\,ds + \int_{a+\alpha_2}^{1-\alpha_2} F^{-1}(s)\,ds\right]$$

Cet estimateur s'écrit :

$$T(F_n) = \frac{1}{n\left(1-2(\alpha_1+\alpha_2)\right)}\left(T_{na\alpha_1} + T_{na\alpha_2}\right)$$

avec :

$$\begin{cases} T_{na\alpha_1} = \displaystyle\sum_{k=[n\alpha_1]+2}^{[n(a-\alpha_1)]} X_{(k)} + \left(1 - n\alpha_1 + [n\alpha_1]\right) X_{([n\alpha_1]+1)} \\ \qquad\qquad + \left(n(a-\alpha_1) - [n(a-\alpha_1)]\right) X_{([n(a-\alpha_1)]+1)} \\ T_{na\alpha_2} = \displaystyle\sum_{k=[n(a+\alpha_2)]+2}^{n-[n\alpha_2]-1} X_{(k)} + \left(1 - n(a+\alpha_2) + [n(a+\alpha_2)]\right) X_{([n(a+\alpha_2)]+1)} \\ \qquad\qquad + \left(1 - n\alpha_2 + [n\alpha_2]\right)\right) X_{(n-[n\alpha_2])} \end{cases}$$

On se ramène au cas précédent en faisant tendre a vers 1. *Cette moyenne α-tronquée généralisée s'interprète de la façon suivante : dans un n-échantillon ordonné, une proportion* $2(\alpha_1 + \alpha_2)$ *des données est enlevée en éliminant les* $[n\alpha_1]$ *plus petites, les* $[n\alpha_2]$ *plus grandes, les* $[n\alpha_1]$ *plus proches du a-quantile tout en lui étant inférieures, et les* $[n\alpha_2]$ *plus proches du a-quantile tout en lui étant supérieures. Nous aurons besoin pour l'un des estimateurs des taux de transition du modèle de Usher de cette troncature en plusieurs points, ce qui jusitifie qu'on s'y intéresse ici.*

3.1.4 Lemmes

Le calcul de la fonction d'influence des L-estimateurs dans le cas unidimensionnel repose sur le lemme suivant. Nous précisons la preuve de ce lemme, puis donnons la preuve d'un second lemme qui nous sera utile pour calculer les fonctions d'influence dans le cas multidimensionnel.

Lemme 1 *: Soit* F *une loi de densité* f *par rapport à la mesure de Lebesgue et strictement croissante sur le support de* f*. Soit* x *un réel. On pose, pour* $s \in \mathbb{R}$ *et pour* $\epsilon > 0$ *:*

$$F_\epsilon(s) = (1-\epsilon)F(s) + \epsilon \mathbf{1}_{s \geq x}$$

Alors, pour $a \in]0;1[$*, la fonction* $\psi_a : \epsilon \mapsto F_\epsilon^{-1}(a)$*, est dérivable en* $\epsilon = 0$*. Pour* $a \in]0;1[\backslash \{F(x)\}$*, la dérivée de* ψ_a *en* 0 *est égale à :*

$$\psi_a^{'}(0) = \frac{a - \mathbf{1}_{F^{-1}(a) \geq x}}{f\left(F^{-1}(a)\right)} \tag{3.7}$$

Pour $a = F(x)$ *:* $\psi_{F(x)}^{'}(0) = 0$*.*

Preuve : La démonstration se fait en plusieurs étapes :
- on montre que pour tout $\epsilon > 0$:

$$[F_\epsilon(x) - \epsilon, F_\epsilon(x)[\subset [F(x) - \epsilon; F(x) + \epsilon[\,;$$

- on en déduit que, pour $a \in]0;1] \setminus \{F(x)\}$ et pour ϵ assez petit on a :

$$F_\epsilon(F_\epsilon^{-1}(a)) = a\,; \tag{3.8}$$

- pour $a \in]0,1[\setminus \{F(x)\}$, on dérive par rapport à ϵ l'égalité (3.8) suivant que $a < F(x)$ ou que $a > F(x)$;
- on traite enfin le cas où $a = F(x)$.

Étape 1 : F_ϵ est continue sur $\mathbb{R} \setminus \{x\}$, strictement croissante et vaut :

$$\begin{aligned} F_\epsilon(s) &= (1-\epsilon)F(s) && \text{pour } s < x \\ &= (1-\epsilon)F(s) + \epsilon && \text{pour } s \geq x \end{aligned}$$

De plus, F_ϵ possède une limite à droite et à gauche en x qui valent :

$$\lim_{s \to x, s<x} F_\epsilon(s) = F_\epsilon(x) - \epsilon$$
$$\lim_{s \to x, s \geq x} F_\epsilon(s) = F_\epsilon(x)$$

De façon générale, si G est une loi sur $\mathbb{R}$, pour tout $b \in]0;1]$:

$$G[G^{-1}(b)] \geq b$$

avec égalité si G est continue. F_ϵ n'est pas continue sur $\mathbb{R}$, donc elle ne vérifie pas l'égalité (3.8) sur $]0;1]$. Néanmoins, elle est continue sur $]-\infty; x[$ à valeurs dans $[0; F_\epsilon(x) - \epsilon[$, et sur $[x; +\infty[$ à valeurs dans $[F_\epsilon(x); 1]$. F_ϵ vérifie donc l'équation (3.8) pour $a \in]0;1] \setminus [F_\epsilon(x) - \epsilon, F_\epsilon(x)[$. Pour tout $b \in [F_\epsilon(x) - \epsilon, F_\epsilon(x)[$:

$$F_\epsilon^{-1}(b) = x$$

et donc :

$$F_\epsilon(F_\epsilon^{-1}(b)) = F_\epsilon(x)$$

De plus, comme $F_\epsilon(x) = F(x) + \epsilon(1 - F(x))$, on a : $F(x) \in [F_\epsilon(x) - \epsilon, F_\epsilon(x)]$, et :

$$[F_\epsilon(x) - \epsilon, F_\epsilon(x)[\subset [F(x) - \epsilon; F(x) + \epsilon[$$

Étape 2 : soit $a \in]0;1]$ et $a \neq F(x)$. Pour tout $\epsilon < |F(x) - a|$, on a :

$$a \notin [F_\epsilon(x) - \epsilon, F_\epsilon(x)[$$

Ainsi, pour tout $\epsilon < |F(x) - a|$:

$$F_\epsilon(F_\epsilon^{-1}(a)) = a$$

Étape 3 : si, maintenant, $a > F(x)$, pour tout $\epsilon < |F(x) - a|$, $F_\epsilon^{-1}(a) > x$ et l'équation précédente s'écrit :

$$(1-\epsilon)F(\psi_a(\epsilon)) + \epsilon = a$$

On en tire l'expression de ψ_a sur $[0; |F(x) - a|[$, F étant bijective :

$$\psi_a(\epsilon) = F^{-1}\left(\frac{a-\epsilon}{1-\epsilon}\right)$$

ψ_a est donc dérivable en $\epsilon = 0$, de dérivée en 0 égale à :

$$\psi_a^{'}(0) = \frac{a-1}{f(F^{-1}(a))}$$

Si $a < F(x)$, on démontre de même que ψ_a est dérivable en 0, dont la dérivée en 0 vaut :

$$\psi_a^{'}(0) = \frac{a}{f(F^{-1}(a))}$$

En résumé, ψ_a est dérivable en 0, pour tout a de $]0;1] \setminus F(x)$, et :

$$\begin{aligned} \psi_a^{'}(0) &= \frac{a - \mathbf{1}_{a>F(x)}}{f(F^{-1}(a))} \\ &= \frac{a - \mathbf{1}_{F^{-1}(a)\geq x}}{f(F^{-1}(a))} \end{aligned}$$

car si $a \neq F(x) : (a > F(x)) \Leftrightarrow (F^{-1}(a) \geq x\ )$.

Étape 4 : pour tout $\epsilon \geq 0$, $F_\epsilon^{-1}(F(x)) = x$. Donc, $\psi_{F(x)}$ est une fonction constante égale à x. Sa dérivée en 0 est donc nulle. ∎

Remarque 2 *: Dans le cas où $F(x) = 1$, l'équation (3.7) est vérifiée pour $a = F(x)$. En effet, pour tout $\epsilon > 0$, $F_\epsilon(x) = F(x) + \epsilon(1 - F(x))$, et pour tout $b \in\ [F_\epsilon(x) -\ \epsilon, F_\epsilon(x)[$:*

$$F_\epsilon(F_\epsilon^{-1}(b)) = F_\epsilon(x)$$

Donc, si $F(x) = 1$, $F_\epsilon(x) = F(x)$ pour tout $\epsilon > 0$, et :

$$F_\epsilon\left[F_\epsilon^{-1}\left(F(x)\right)\right] = F(x)$$

En appliquant l'équation (3.7) en $a = F(x)$, on retrouve bien : $\psi_{F(x)}^{'}(0) = 0$.

Par contre, si $F(x) < 1$, l'équation (3.7) est fausse pour $a = F(x)$. En effet, dans ce cas, $F(x) < F_\epsilon(x)$ pour tout $\epsilon > 0$, et :

$$F_\epsilon\left[F_\epsilon^{-1}\left(F(x)\right)\right] \neq\ F(x)$$

pour tout $\epsilon > 0$. D'ailleurs, si on applique l'équation (3.7) au point $a = F(x)$, on ne retrouve pas la valeur de $\psi_{F(x)}^{'}(0)$.

Le lemme précédent peut s'étendre au lemme suivant :

Lemme 2 *: Soit F une loi de densité f et strictement croissante sur le support de f. Soit x un réel. On pose, pour $s \in\ \mathbb{R}$ et pour $\epsilon > 0$:*

$$F_\epsilon(s) = (1-\epsilon)F(s) + \epsilon\mathbf{1}_{s\geq x}$$

Soit $a \in]0;1[$, et une fonction A de $\mathbb{R}^+$ dans $[0;1]$, dérivable en 0 et telle que $A(0) = a$. Alors, la fonction $\varphi_a : \epsilon \mapsto F_\epsilon^{-1}(A(\epsilon))$, est dérivable en $\epsilon = 0$. Pour $a \in]0;1[\backslash\{F(x)\}$, la dérivée de φ_a en 0 est égale à :

$$\varphi_a^{'}(0) = \frac{a + A^{'}(0) - \mathbf{1}_{F^{-1}(a) \geq x}}{f\left(F^{-1}(a)\right)} \tag{3.9}$$

Pour $a = F(x)$ et si A est telle que, pour tout $\epsilon > 0$:

$$-\epsilon a \leq A(\epsilon) - a < \epsilon(1 - a) \tag{3.10}$$

alors : $\varphi_{F(x)}^{'}(0) = 0$.

Preuve : Le raisonnement est le même que pour la preuve du lemme précédent. On montre que :

- pour $a \in]0;1] \setminus \{F(x)\}$, on a pour ϵ assez petit :

$$F_\epsilon[F_\epsilon^{-1}(A(\epsilon))] = A(\epsilon)\,; \tag{3.11}$$

- pour $a \in]0,1[\backslash\{F(x)\}$, on en déduit l'expression de la dérivée de φ ;
- on traite le cas où $a = F(x)$.

Étape 1 : Soit $a \in]0;1]$ et $a \neq F(x)$. Comme A est continue en 0 et $A(0) = a$ alors :

$$\exists E,\, \forall \epsilon < E,\, |a - A(\epsilon)| < \frac{|F(x) - a|}{2}$$

Ainsi, si $E^{'} = \min\left(E, \frac{|F(x) - a|}{2}\right)$, pour tout $\epsilon < E^{'}$, on a :

$$\begin{aligned} |A(\epsilon) - F(x)| &> |a - F(x)| - |a - A(\epsilon)| \\ &> \frac{|F(x) - a|}{2} \\ &> \epsilon \end{aligned}$$

Donc, pour tout $\epsilon < E^{'}$:

$$A(\epsilon) \notin [F(x) - \epsilon, F(x) + \epsilon]$$

donc, comme $[F_\epsilon(x) - \epsilon, F_\epsilon(x)[\subset [F(x) - \epsilon; F(x) + \epsilon[$ (voir la démonstration du lemme précédent) :

$$A(\epsilon) \notin [F_\epsilon(x) - \epsilon, F_\epsilon(x)[$$

Ainsi :

$$F_\epsilon[F_\epsilon^{-1}(A(\epsilon))] = A(\epsilon)$$

Étape 2 : si $a > F(x)$, pour tout $\epsilon < E^{'}$,$A(\epsilon) > F_\epsilon(x)$ et $F_\epsilon^{-1}(A(\epsilon)) > x$. L'équation précédente s'écrit alors :

$$(1-\epsilon)F(\varphi_a(\epsilon)) + \epsilon = A(\epsilon)$$

φ_a est donc dérivable en $\epsilon = 0$, de dérivée en 0 égale à :

$$\varphi_a^{'}(0) = \frac{a + A^{'}(0) - 1}{f(F^{-1}(a))}$$

Si $a < F(x)$, on démontre de même que φ_a est dérivable en 0, dont la dérivée en 0 vaut :

$$\varphi_a^{'}(0) = \frac{a + A^{'}(0)}{f(F^{-1}(a))}$$

En résumé, φ_a est dérivable en 0, pour tout a de $]0;1] \setminus F(x)$, et :

$$\begin{aligned} \varphi_a^{'}(0) &= \frac{a + A^{'}(0) - \mathbf{1}_{a>F(x)}}{f(F^{-1}(a))} \\ &= \frac{a + A^{'}(0) - \mathbf{1}_{F^{-1}(a)\geq x}}{f(F^{-1}(a))} \end{aligned}$$

car si $a \neq F(x)$: $(a > F(x)) \Leftrightarrow (F^{-1}(a) \geq x\ )$.

Étape 3 : si $a = F(x)$ et A est telle que, pour tout $\epsilon > 0$:

$$-\epsilon F(x) \leq A(\epsilon) - F(x) < \epsilon(1 - F(x))$$

comme :

$$A(\epsilon) - F_\epsilon(x) = (A(\epsilon) - F(x)) - \epsilon(1 - F(x))$$

alors :

$$A(\epsilon) \in [F_\epsilon(x) - \epsilon, F_\epsilon(x)[$$

Donc, pour tout $\epsilon > 0$, $F_\epsilon^{-1}(A(\epsilon)) = x$. Donc, $\varphi_{F(x)}$ est une fonction constante égale à x. Sa dérivée est donc nulle. ∎

Ces lemmes seront utilisés dans le calcul de la fonction d'influence des estimateurs des taux de transition du modèle de Usher.

3.2 Robustesse dans le modèle de Usher

3.2.1 Cadre de l'étude

Dans ce paragraphe, on s'intéresse à l'étude de la robustesse des estimateurs des taux de transition dans le modèle de Usher, afin de traiter les données qui ne respectent pas l'hypothèse de Usher, d'une part, et pour limiter l'impact des erreurs de mesure sur la dynamique de la population, d'autre part. Ces deux questions se rapportent aux individus qui restent vivants. On s'intéresse donc aux taux de transition conditionnellement au fait que l'individu reste vivant. Par ailleurs, les erreurs de classification viennent des erreurs de mesure portant sur la taille des individus, modélisées par des variables continues. On spécifie donc ici le modèle statistique sur ces variables et on étudie la robustesse des estimateurs des taux de transition dans ce modèle, en comparant leur sensibilité.

Modèle statistique

On considère une classe $C = [l_0, l_1[$ et sa classe supérieure $C' = [l_1, l_2[$. Soient X la taille à un temps t d'un individu de la classe C qui reste vivant entre t et $t+1$, et ΔX son accroissement en taille entre ces deux pas de temps. On suppose que :

- X suit une loi de densité f et de fonction de répartition F. La fonction f est définie sur $[l_0, l_1[$;
- la loi de $\Delta X|X$ est de densité $g(.|X)$ et de fonction de répartition G_X. Pour chaque taille x, le support de la fonction $y \mapsto g(.|x)$ est égal à $[-x; +\infty[$.

La loi du couple $(X, \Delta X)$, notée H, a pour densité :

$$h(x, y) = g(y|x)f(x)$$

Définitions des taux de transition

On s'intéresse au taux de transition $q^\bullet$ correspondant à la probabilité qu'un individu de la classe C passe dans la classe supérieure C' sachant qu'il reste vivant. On considère ici deux questions relatives à $q^\bullet$ qui peuvent être traitées dans le cadre de la robustesse.

La première de ces questions est : comment gérer les données qui violent l'hypothèse d'Usher ? Deux possibilités s'offrent pour traiter ces données (les éliminer ou les lisser), ce qui débouche sur deux versions de $q^\bullet$. La probabilité de transition correspondant à l'estimateur classique, noté $q^\bullet_{M\,M'}$, est liée à la troncature sur la variable taille continue X. Il est défini par la probabilité

conditionnelle qu'un individu de la classe C au temps t passe dans la classe supérieure sachant qu'au temps $t+1$ la taille de l'individu est plus petit qu'un réel $M^{'}$ et plus grand qu'un réel M. Les réels M et $M^{'}$ vérifient les conditions suivantes : $0 \leq M^{'} \leq l_0$ et $l_1 < M$. L'élimination des individus qui ne respectent pas l'hypothèse de Usher correspond à $M^{'} = l_0$ et $M = l_2$. On compare l'estimateur classique à l'estimateur qui ne tronque pas les données. Le taux de transition qui correspond à ce dernier est défini par la probabilité qu'un individu de la classe C au temps t est plus grand en taille que l_1 au temps $t+1$. Il conserve ainsi les individus dont la taille est supérieure à l_2 ce qui revient à lisser les données. Cette probabilité, notée $q^{\bullet}_{\infty}$, est la limite de $q^{\bullet}_{M\,M'}$ quand M est égal à zéro et $M^{'}$ tend vers l'infini. Ces probabilités s'écrivent :

$$q^{\bullet}_{MM'} = \Pr(l_1 \leq X + \Delta X | M' \leq X + \Delta X < M) \quad (3.12)$$

$$q^{\bullet}_{\infty} = \Pr(l_1 \leq X + \Delta X) \quad (3.13)$$

La seconde question est : dès lors que les données vérifient l'hypothèse d'Usher, comment minimiser l'impact des erreurs de mesure de la taille et de l'accroissement en taille ? L'impact des erreurs de mesure peut être réduit en tronquant les valeurs extrêmes des tailles et/ou des accroissements en taille. Sachant que $X + \Delta X \in [l_0; l_2[$, on envisage deux façons de tronquer les données, ce qui débouche sur deux versions des taux de transition :

1. Soit $0 < \alpha < 1$, on fait une α-troncature sur la loi de ΔX : on élimine les couples $(X, \Delta X)$ tel que :

   $$\Delta X \geq l_2 - l_0 - \delta$$

   où δ est déterminé de telle sorte à éliminer une proportion α d'individus ;

2. Soit $0 < \alpha^{'} < 1$, on fait une $\alpha^{'}$-troncature sur les intervalles $[l_0; l_1[$ et $[l_1; l_2[$: on élimine les couples $(X, \Delta X)$ tels que X ou $X + \Delta X$ appartiennent aux bords des intervalles, soit, si on note $\delta^{'} > 0$:

   $$\begin{cases} X \in [l_0; l_0 + \delta^{'}] \cup [l_1 - \delta^{'}; l_1] \\ \text{ou} \\ X + \Delta X \in [l_0; l_0 + \delta^{'}] \cup [l_1 - \delta^{'}, l_1 + \delta^{'}] \cup [l_2 - \delta^{'}; l_2[\end{cases}$$

   où $\delta^{'}$ est déterminé de manière à enlever une proportion $\alpha^{'}$ des observations.

On note respectivement ces paramètres $q_1^\bullet$ et $q_2^\bullet$ et on pose :

$$\begin{aligned} I_{\delta'} &=]l_0+\delta^{'}; l_1-\delta^{'}[\\ J_{\delta'} &=]l_0+\delta^{'}; l_1-\delta^{'}[\cup]l_1+\delta^{'}; l_2-\delta^{'}[\end{aligned}$$

Ils s'expriment de la façon suivante :

$$q_1^\bullet = \Pr(l_1 \leq X+\Delta X | l_0 \leq X+\Delta X < l_2 \,\text{et}\, \Delta X < l_2-l_0-\delta) \quad (3.14)$$

$$q_2^\bullet = \Pr(l_1 \leq X+\Delta X | X \in I_{\delta'} \,\text{et}\, X+\Delta X \in J_{\delta'}) \quad (3.15)$$

Le paramètre $q_2^\bullet$ est une probabilité conditionnelle aussi au fait que $X+\Delta X \in [l_0; l_2[$, mais cette condition est redondante car $J_{\delta'} \subset [l_0; l_2[$. De plus, δ et $\delta^{'}$ sont déterminés respectivement par les équations :

$$\Pr(\Delta X < l_2-l_0-\delta | l_0 < X+\Delta X < l_2) = 1-\alpha \quad (3.16)$$

$$\Pr(X \in I_{\delta'} \,\text{et}\, X+\Delta X \in J_{\delta'} | l_0 < X+\Delta X < l_2) = 1-\alpha^{'} \quad (3.17)$$

La robustesse des estimateurs de $q_1^\bullet$ et $q_2^\bullet$ sera comparée à celle de l'estimateur classique où aucune donnée extrême n'est tronquée. Partant du n-échantillon incluant les données qui violent l'hypothèse d'Usher, le taux de transition correspondant s'identifie à $q_{MM'}^\bullet$ avec $M'=l_0$ et $M=l_2$.

3.2.2 Robustesse des estimateurs des probabilités de transition

Dans cette partie, on étudie la robustesse des estimateurs des probabilités de transition proposés. On exprime tout d'abord les fonctionnelles qui les engendrent. On calcule ensuite leur fonction d'influence afin de calculer leur sensibilité et leur variance asymptotique.

Estimateurs des probabilités de transition

Étant donné un n-échantillon $(X_1, \ldots, X_n)$ de loi H, les estimateurs empiriques de $q_{M\,M'}^\bullet$, $q_\infty^\bullet$, $q_1^\bullet$ et $q_2^\bullet$ sont respectivement :

$$\begin{aligned} \hat{q}_{MM'}^\bullet &= \frac{\sum_{k=1}^n \mathbf{1}_{l_1<X_k+\Delta X_k<M}}{\sum_{k=1}^n \mathbf{1}_{M'\leq X_k+\Delta X_k<M}} \\ \hat{q}_\infty^\bullet &= \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{l_1<X_k+\Delta X_k} \\ \hat{q}_1^\bullet &= \frac{\sum_{k=1}^n \mathbf{1}_{l_1<X_k+\Delta X_k<l_2,\, \Delta X_k<l_2-l_0-\delta}}{\sum_{k=1}^n \mathbf{1}_{l_0<X_k+\Delta X_k<l_2,\, \Delta X_k<l_2-l_0-\delta}} \\ \hat{q}_2^\bullet &= \frac{\sum_{k=1}^n \mathbf{1}_{X_k+\Delta X_k\in J_{\delta'}^1,\, X_k\in I_{\delta'}}}{\sum_{k=1}^n \mathbf{1}_{X_k+\Delta X_k\in J_{\delta'},\, X_k\in I_{\delta'}}} \end{aligned}$$

où $J^1_{\delta'} =]l_1 + \delta' ; l_2 - \delta' [$.

On exprime alors les fonctionnelles qui engendrent ces estimateurs. On note les ensembles (voir la figure 3.1) :

$$\begin{aligned} \mathcal{D} &= \{(x,y); l_0 \leq x < l_1, x+y \geq 0\}, \\ \mathcal{D}_2 &= \{(x,y) \in \mathcal{D}; x+y \geq l_1\} \end{aligned}$$

et :

$$\begin{aligned} \mathcal{D}_M &= \{(x,y) \in \mathcal{D}; x+y < M\} \\ \mathcal{D}_{M'} &= \{(x,y) \in \mathcal{D}; x+y \geq M'\} \\ \mathcal{D}_{MM'} &= \mathcal{D}_M \cap \mathcal{D}_{M'} \end{aligned}$$

On note aussi $\mathcal{D}_1$ le complémentaire de $\mathcal{D}_2$ dans $\mathcal{D}$, défini par :

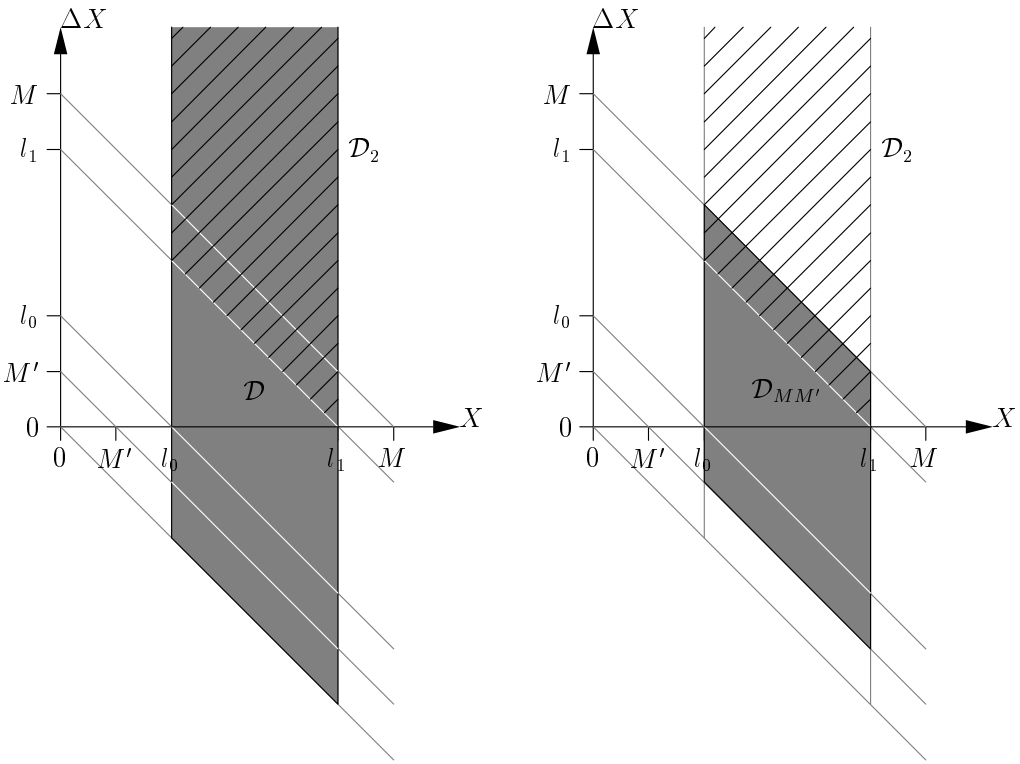


FIG. 3.1 –

$$\mathcal{D}_1 = \{(x,y) \in \mathcal{D}; x+y < l_1\}$$

Dans le cas particulier où $M = l_0$ et $M^{'} = l_2$, on note (voir la figure 3.2) :

$$\begin{aligned}\tilde{\mathcal{D}} &= \mathcal{D}_{l_2\, l_0}\\ \tilde{\mathcal{D}}_1 &= \mathcal{D}_1 \cap \tilde{\mathcal{D}}\\ \tilde{\mathcal{D}}_2 &= \mathcal{D}_2 \cap \tilde{\mathcal{D}}\end{aligned}$$

On note de plus les ensembles $\mathcal{D}_\delta$ et $\mathcal{D}_{\delta'}$ définis par (voir les figures 3.3 et 3.4) :

$$\begin{aligned}\mathcal{D}_\delta &= \{(x,y) \in \mathcal{D}; y < l_2 - l_0 - \delta\}\\ \mathcal{D}_{\delta'} &= \{(x,y) \in \mathcal{D}; x \in I_{\delta'}, x + y \in J_{\delta'}\}\end{aligned}$$

et déterminés par les équations :

$$\begin{aligned}\iint_{\mathcal{D}_\delta \cap \tilde{\mathcal{D}}} h(x,y)\mathrm{d}x\mathrm{d}y &= (1-\alpha)\iint_{\tilde{\mathcal{D}}} h(x,y)\mathrm{d}x\mathrm{d}y\\ \iint_{\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}} h(x,y)\mathrm{d}x\mathrm{d}y &= (1-\alpha^{'})\iint_{\tilde{\mathcal{D}}} h(x,y)\mathrm{d}x\mathrm{d}y\end{aligned}$$

Soient T et $T_{M\,M'}$ les fonctionnelles qui engendrent respectivement les estimateurs empiriques de $q^\bullet_\infty$ et de $q^\bullet_{M\,M'}$ d'une part, et T_1 et T_2 les fonctionnelles qui engendrent respectivement les estimateurs empiriques de $q^\bullet_1$ et $q^\bullet_2$ d'autre part, définis par les équations :

$$T_{MM'}(H) = \frac{\iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} h(x,y)\mathrm{d}x\mathrm{d}y}{\iint_{\mathcal{D}_{MM'}} h(x,y)\mathrm{d}x\mathrm{d}y} \tag{3.18}$$

$$T(H) = \iint_{\mathcal{D}_2} h(x,y)\mathrm{d}y\mathrm{d}x \tag{3.19}$$

$$T_1(H) = \frac{\iint_{\mathcal{D}_\delta \cap \tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y}{\iint_{\mathcal{D}_\delta \cap \tilde{\mathcal{D}}} h(x,y)\mathrm{d}x\mathrm{d}y} \tag{3.20}$$

$$T_2(H) = \frac{\iint_{\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y}{\iint_{\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}} h_i(x,y)\mathrm{d}x\mathrm{d}y} \tag{3.21}$$

On peut relier les fonctionnelles entre eux. Pour cela, on pose :

$$\begin{aligned}\alpha_{MM'} &= \iint_{\mathcal{D}\setminus\mathcal{D}_{MM'}} h(x,y)\mathrm{d}x\mathrm{d}y\\ &= 1 - \iint_{\mathcal{D}_{MM'}} h(x,y)\mathrm{d}x\mathrm{d}y\end{aligned} \tag{3.22}$$

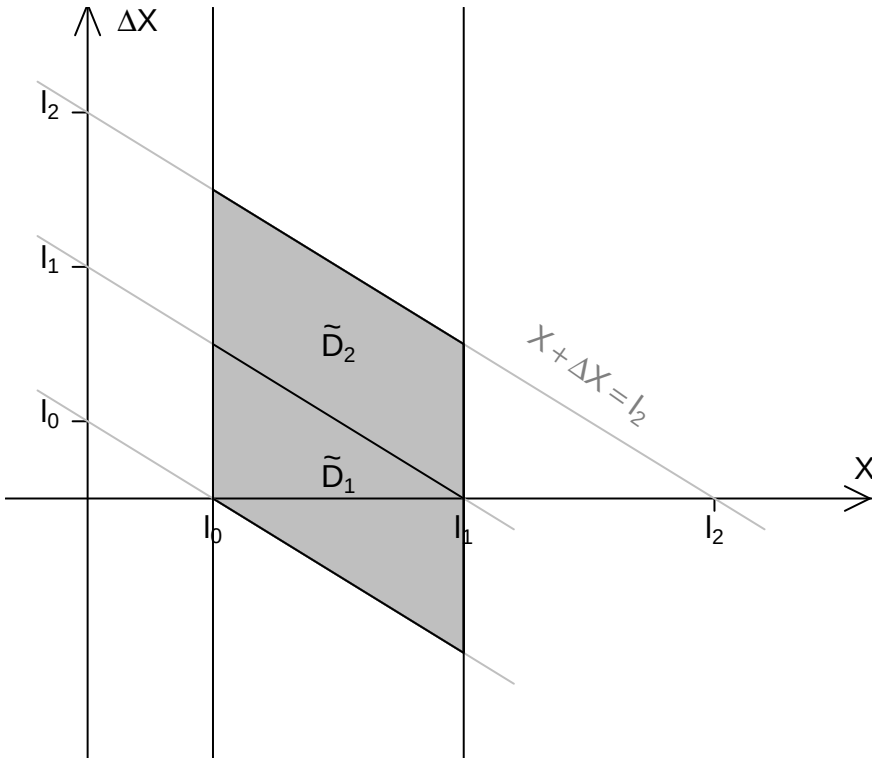


FIG. 3.2 –

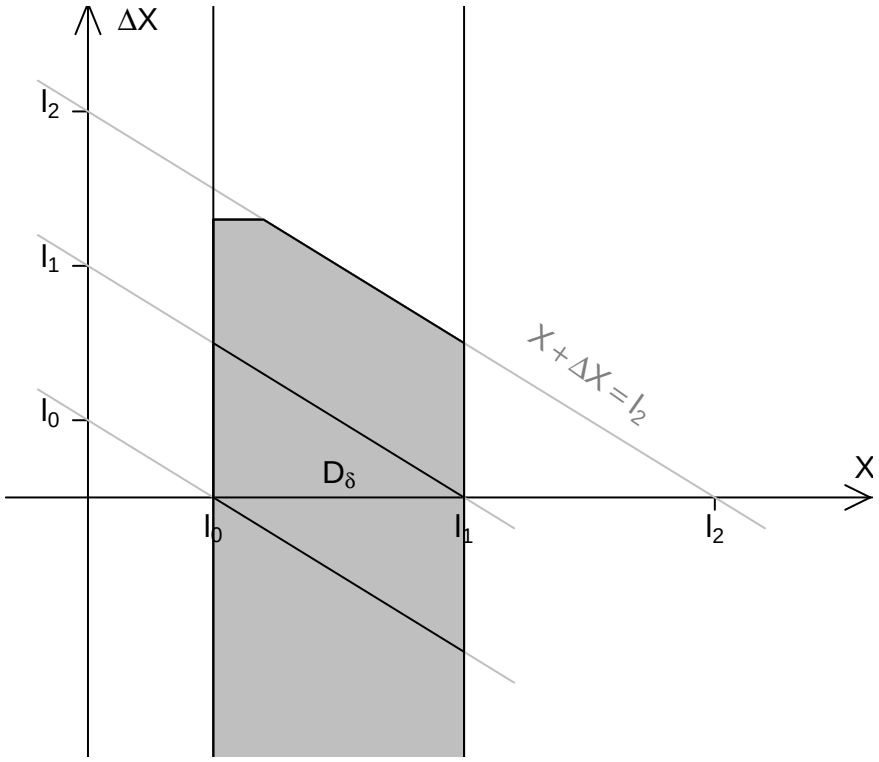


FIG. 3.3 – Troncature sur les accroissements

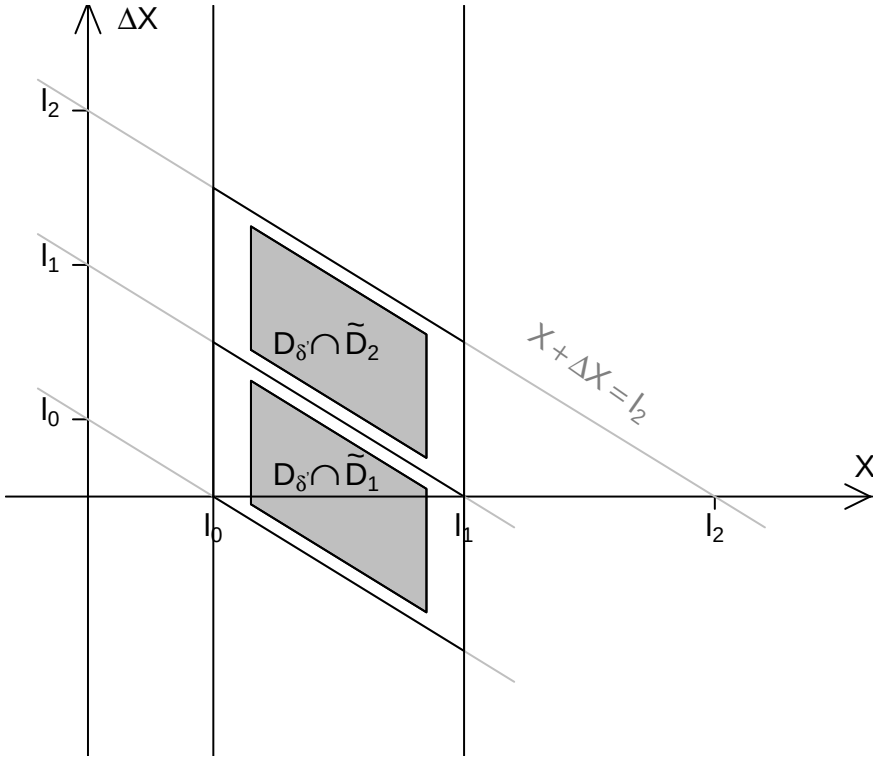


FIG. 3.4 – Troncature sur les bornes des classes. Le domaine grisé correspond à l'ensemble $\mathcal{D}_{\delta'}$.

et :

$$\begin{aligned}\alpha_M &= 1 - \iint_{\mathcal{D}_M} h(x,y)\mathrm{d}x\mathrm{d}y \\ \alpha_{M'} &= 1 - \iint_{\mathcal{D}_{M'}} h(x,y)\mathrm{d}x\mathrm{d}y\end{aligned}$$

Le réel $\alpha_{MM'}$ vérifie : $0 \leq \alpha_{MM'} \leq 1$, et : $\alpha_{MM'} = \alpha_M + \alpha_{M'}$. Le cas où $\alpha_{MM'} = 1$ correspond au cas où tous les individus ont leur taille au temps $t+1$ plus grand que M ou entre zéro et M'. Par la suite, on ne se placera pas dans ce cas et on supposera que : $0 \leq \alpha_{MM'} < 1$. D'autre part, on note que $\alpha_{M\,M'}$ dépend de la loi H alors que α et $\alpha^{'}$ sont fixés au départ. Ce sont les ensembles $\mathcal{D}_\delta$ et $\mathcal{D}_{\delta'}$ qui dépendent de H, et de α et $\alpha^{'}$.

On a alors les relations suivantes :

$$T_{M\,M'}(H) = \frac{T(H) - \alpha_M}{1 - \alpha_{MM'}} \tag{3.23}$$

$$T_1(H) = \frac{T(H) - \alpha_{l_2} - \alpha(1 - \alpha_{l_2\,l_0})}{(1-\alpha)(1-\alpha_{l_2\,l_0})} \tag{3.24}$$

$$T_2(H) = \frac{T(H) - \alpha_{l_2} - \beta^{'}(1 - \alpha_{l_2\,l_0})}{(1-\alpha^{'})(1-\alpha_{l_2\,l_0})} \tag{3.25}$$

avec : $\beta^{'} = \iint_{\tilde{\mathcal{D}}_2 \setminus \left(\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}_2\right)} h(x,y)\mathrm{d}x\mathrm{d}y.$

La relation entre T et $T_{M\,M'}$ est détaillée dans l'annexe C. D'autre part :

$$(1 - \alpha_{l_2\,l_0})T_{l_2\,l_0}(H) = \iint_{\mathcal{D}_\delta \cap \tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y + \iint_{\tilde{\mathcal{D}} \setminus \left(\mathcal{D}_\delta \cap \tilde{\mathcal{D}}\right)} h(x,y)\mathrm{d}x\mathrm{d}y$$

car : $\tilde{\mathcal{D}} \setminus \left(\mathcal{D}_\delta \cap \tilde{\mathcal{D}}\right) \subset \tilde{\mathcal{D}}_2$. Alors :

$$(1 - \alpha_{l_2\,l_0})T_{l_2\,l_0}(H) = (1-\alpha)(1-\alpha_{l_2\,l_0})T_1(H) + \alpha(1 - \alpha_{l_2\,l_0})$$

et :

$$T_{l_2\,l_0}(H) = (1-\alpha)T_1(H) + \alpha \tag{3.26}$$

On en déduit la relation entre $T_1(H)$ et $T(H)$ par la relation (3.23). On fait de même pour $T_2(H)$, à la différence que l'ensemble $\tilde{\mathcal{D}} \setminus \left(\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}\right)$ n'est pas inclus dans $\tilde{\mathcal{D}}_2$. La relation entre T_2 et $T_{l_2\,l_0}$ est donnée par l'équation :

$$T_{l_2\,l_0}(H) = (1-\alpha^{'})T_2(H) + \beta^{'} \tag{3.27}$$

Remarque 3 *: On a en particulier :*

$$T(H) \geq \alpha_M$$

De plus, comme $T_1(H) \geq 0$ *et* $T_2(H) \geq 0$*, on doit choisir* α *et* $\alpha^{'}$ *de telles sortes que :* $T_{l_2\, l_0}(H) \geq \alpha$ *et* $T_{l_2\, l_0}(H) \geq \alpha^{'}$ *ou tel que :*

$$\frac{T(H) - \alpha_{l_2}}{1 - \alpha_{l_2\, l_0}} \geq \alpha$$
$$\frac{T(H) - \alpha_{l_2}}{1 - \alpha_{l_2\, l_0}} \geq \alpha^{'}$$

Calcul de la fonction d'influence

Dans cette partie, on calcule les fonctions d'influence des fonctionnelles $T_{MM'}$ et T d'une part et de T_1 et T_2 d'autre part. On note (u, v) un couple de $\mathcal{D}$.

L'expression de la fonction d'influence de $T_{MM'}$ au point H dans la direction (u, v), $IC_{T_{MM'},H}(u, v)$, est donnée par la proposition suivante :

Proposition 2 *La fonction d'influence de l'estimateur tronqué généré par la fonctionnelle* $T_{MM'}$*, au point de la distribution* H*, dans la direction du point* (u, v) *dans* $\mathcal{D}$*,* $IC_{T_{MM'},H}$*, est égale à :*

$$IC_{T_{MM'},H}((u,v)) = \begin{cases} -\dfrac{T(H) - \alpha_M}{(1 - \alpha_{MM'})^2} & \text{pour } (u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_1 \\ \dfrac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{MM'})^2} & \text{pour } (u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2 \\ 0 & \text{pour } (u,v) \in \mathcal{D} \setminus \mathcal{D}_{MM'} \end{cases}$$

où les domaines $\mathcal{D}$*,* $\mathcal{D}_1$*,* $\mathcal{D}_2$ *et* $\mathcal{D}_{MM'}$*, et les proportions* $\alpha_{MM'}$*,* α_M *et* $\alpha_{M'}$ *sont définis dans le précédent paragraphe.*

La fonctionnelle T étant un cas limite de $T_{MM'}$, on en déduit la fonction d'influence de T au point H dans la direction (u, v), donné par le corollaire suivant :

Corollaire 1 *La fonction d'influence de l'estimateur non tronqué généré par la fonctionnelle* T*, au point de la distribution* H*, dans la direction du point* (u, v) *dans* $\mathcal{D}$*,* $IC_{T,H}$*, est égale à :*

$$IC_{T,H}((u,v)) = \begin{cases} -T(H) & \text{pour } (u,v) \in \mathcal{D}_1 \\ 1 - T(H) & \text{pour } (u,v) \in \mathcal{D}_2 \end{cases}$$

où les domaines $\mathcal{D}$, $\mathcal{D}_1$, $\mathcal{D}_2$ sont définis dans le précédent paragraphe.

Preuve : Les preuves de la proposition 2 et du corollaire 1 sont données dans l'annexe C.

D'autre part, la fonction d'influence de T_1 au point H dans la direction (u, v), est donnée par la proposition suivante :

Proposition 3 *La fonction d'influence de l'estimateur tronqué généré par la fonctionnelle T_1, au point de la distribution H, dans la direction du point (u, v) dans $\mathcal{D}$, $IC_{T_1,H}$, est égale à :*

$$IC_{T_1,H}((u,v)) = \begin{cases} -\dfrac{T(H) - \alpha_{l_2}}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2} & \text{pour } (u,v) \in \tilde{\mathcal{D}}_1 \\ \dfrac{1 - \alpha_{l_0} - T(H)}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2} & \text{pour } (u,v) \in \tilde{\mathcal{D}}_2 \text{ et } v \neq l_2 - l_0 - \delta \\ -\dfrac{T(H) - \alpha_{l_2} - \alpha(1-\alpha_{l_2\,l_0})}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2} & \text{pour } (u,v) \in \tilde{\mathcal{D}}_2 \text{ et } v = l_2 - l_0 - \delta \\ 0 & \text{sinon} \end{cases}$$

Remarque 4 *: si on pose $\alpha = 0$ dans l'expression de $IC_{T_1,H}$, ce qui entraîne $\delta = 0$, on retrouve l'expression de la fonction d'influence de l'estimateur $T_{l_2\,l_0}$.*

Preuve : Le calcul de la fonction d'influence de T_1 se fait selon les étapes suivantes :

- on détermine l'expression de la fonction d'influence de T_1 comme une limite à calculer. On exprime en particulier l'ensemble $\mathcal{D}_\delta$ en fonction de la loi des accroissements en taille ;
- on calcule la fonction d'influence en utilisant les résultats du lemme 2.

Étape 1 : soit G la loi de $\Delta X | X \in C$, de densité g et définie, pour $t \in [-l_0; +\infty[$, par :

$$G(t) = \int_{l_0}^{l_1} G_x(t) f(x) \mathrm{d}x$$

L'ensemble $\mathcal{D}_\delta$ est alors défini par :

$$\mathcal{D}_\delta = \left\{(x,y) \in \mathcal{D}; y < G^{-1}\left(1 - \alpha_{l_2} - \alpha(1 - \alpha_{l_2\,l_0})\right)\right\}$$

ce qui s'écrit aussi, si on pose $\beta = \alpha_{l_2} + \alpha(1 - \alpha_{l_2\,l_0})$:

$$\mathcal{D}_\delta = \left\{(x,y) \in \mathcal{D}; y < G^{-1}\left(1 - \beta\right)\right\}$$

La fonction d'influence de T_1 en H dans la direction (u,v) s'écrit :

$$IC_{T_1,H}((u,v)) = \lim_{\epsilon\to 0}\frac{1}{\epsilon(1-\alpha)}\left[\frac{\iint_{\mathcal{D}_\delta^\epsilon\cap\tilde{\mathcal{D}}_2}(1-\epsilon)h(x,y)\mathrm{d}x\mathrm{d}y+\epsilon\mathbf{1}_{(u,v)\in\mathcal{D}_\delta^\epsilon\cap\tilde{\mathcal{D}}_2}}{\iint_{\tilde{\mathcal{D}}}(1-\epsilon)h(x,y)\mathrm{d}x\mathrm{d}y+\epsilon\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}} - \frac{\iint_{\mathcal{D}_\delta\cap\tilde{\mathcal{D}}_2}h(x,y)\mathrm{d}x\mathrm{d}y}{(1-\alpha_{l_2\,l_0})}\right]$$

avec :

$$\mathcal{D}_\delta^\epsilon = \left\{(x,y)\in\mathcal{D}; y < G_\epsilon^{-1}(1-\beta_\epsilon)\right\}$$

où :

$$\begin{aligned} G_\epsilon(t) &= \int_{l_i}^{l_{i+1}}\left[(1-\epsilon)G_x(t)f(x)+\epsilon\mathbf{1}_{\{u\in[l_0;l_1[,t\geq v\}}\right]\mathrm{d}x \\ &= (1-\epsilon)G(t)+\epsilon\mathbf{1}_{t\geq v} \end{aligned}$$

G_ϵ étant une distribution dont l'inverse est définie, pour $a\in]0;1]$, par :

$$G_\epsilon^{-1}(a) = \inf\{t; G_\epsilon(t)\geq a\}$$

et où :

$$\begin{aligned} 1-\beta_\epsilon &= \iint_{\mathcal{D}_{l_2}}(1-\epsilon)h(x,y)\mathrm{d}x\mathrm{d}y+\epsilon\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} \\ &\quad -\alpha\left(\iint_{\tilde{\mathcal{D}}}(1-\epsilon)h(x,y)\mathrm{d}x\mathrm{d}y+\epsilon\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}\right) \\ &= (1-\epsilon)(1-\alpha_{l_2})+\epsilon\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} \\ &\quad -\alpha\left[(1-\epsilon)(1-\alpha_{l_2\,l_0})+\epsilon\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}\right] \\ &= (1-\epsilon)(1-\beta)+\epsilon(\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}}-\alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}) \end{aligned} \tag{3.28}$$

Le réel $1-\beta_\epsilon$ vérifie, pour tout $\epsilon>0$:

$$-\epsilon(1-\beta)\leq(1-\beta_\epsilon)-(1-\beta)<\epsilon(1-a) \tag{3.29}$$

En effet, d'après l'équation (3.28), :

$$(1-\beta_\epsilon)-(1-\beta) = \epsilon[-(1-\beta)+\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}}-\alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}]$$

et :

$$\begin{aligned} &\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}}-\alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}<1 \\ &\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}}-\alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}>0 \end{aligned}$$

car $\mathcal{D}_{l_2} \supset \tilde{\mathcal{D}}$ et donc $\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} > \mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}$.

D'autre part, on note :

$$\begin{aligned}
\tilde{\Delta}(u,v) &= \mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} \\
\Delta_\epsilon(u,v) &= \mathbf{1}_{(u,v)\in\mathcal{D}_\delta^\epsilon\cap\tilde{\mathcal{D}}_2} \\
I &= \iint_{\mathcal{D}_\delta\cap\tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y \\
I_\epsilon &= \iint_{\mathcal{D}_\delta^\epsilon\cap\tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

Ainsi :

$$\begin{aligned}
IC_{T_1,H}((u,v)) &= \lim_{\epsilon\to 0}\frac{1}{\epsilon(1-\alpha)}\left[\frac{(1-\epsilon)I_\epsilon+\epsilon\Delta_\epsilon(u,v)}{(1-\epsilon)(1-\alpha_{l_2\,l_0})+\epsilon\tilde{\Delta}(u,v)}-\frac{I}{(1-\alpha_{l_2\,l_0})}\right] \\
&= \lim_{\epsilon\to 0}\frac{1}{(1-\alpha)}\left[\frac{(1-\alpha_{l_2\,l_0})\Delta_\epsilon(u,v)-\tilde{\Delta}(u,v)\,I+(1-\epsilon)(1-\alpha_{l_2\,l_0})\frac{I_\epsilon-I}{\epsilon}}{[(1-\epsilon)(1-\alpha_{l_2\,l_0})+\epsilon\tilde{\Delta}(u,v)](1-\alpha_{l_2\,l_0})}\right]
\end{aligned}$$

On va montrer que $\Delta_\epsilon(u,v)$ et que $\dfrac{I_\epsilon - I}{\epsilon}$ ont une limite quand ϵ tend vers zéro. Ainsi :

$$IC_{T_1,H}((u,v)) = \frac{1}{(1-\alpha)(1-\alpha_{l_2\,l_0})}\left[\lim_{\epsilon\to 0}\Delta_\epsilon(u,v)+\lim_{\epsilon\to 0}\frac{I_\epsilon-I}{\epsilon}-\frac{I}{1-\alpha_{l_2\,l_0}}\tilde{\Delta}(u,v)\right]$$

Étape 2 : calculons la limite de $\Delta_\epsilon(u,v)$ quand ϵ tend vers zéro. D'après le lemme 2 et l'inégalité (3.29), la fonction $\epsilon \mapsto G_\epsilon^{-1}(1-\beta_\epsilon)$ est continue en $\epsilon = 0$. Ainsi :

$$G_\epsilon^{-1}(1-\beta_\epsilon) \xrightarrow{\epsilon\to 0} G^{-1}(1-\beta)$$

Si $(u,v) \in \mathcal{D}_\delta \cap \tilde{\mathcal{D}}_2$, $v < G^{-1}(1-\beta)$, par continuité, on a :

$$\exists E_1 > 0, \forall \epsilon < E_1, |G_\epsilon^{-1}(1-\beta_\epsilon) - G^{-1}(1-\beta)| < G^{-1}(1-\beta) - v$$

Ainsi, pour tout $\epsilon < E_1$, $v < G_\epsilon^{-1}(1-\beta_\epsilon)$ donc $(u,v) \in \mathcal{D}_\delta^\epsilon \cap \tilde{\mathcal{D}}_2$. En résumé :

$$\exists E_1 > 0, \forall \epsilon < E_1, \Delta_\epsilon(u,v) = 1$$

De même, si $(u,v) \notin \mathcal{D}_\delta \cap \tilde{\mathcal{D}}_2$ et si $v \neq G^{-1}(1-\beta)$:

$$\exists E_2 > 0, \forall \epsilon < E_2, \Delta_\epsilon(u,v) = 0$$

Dans le cas où $v = G^{-1}(1-\beta)$, on a :

$$G_\epsilon(v) = (1-\epsilon)(1-\beta) + \epsilon$$

D'après l'équation (3.28), ceci s'écrit aussi :

$$G_\epsilon(v) = (1-\beta_\epsilon) + \epsilon(1 - \mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} + \alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}})$$

On en déduit que : $1-\beta_\epsilon < G_\epsilon(v)$. Ainsi $G_\epsilon^{-1}(1-\beta_\epsilon) \leq v$ et $v \notin \mathcal{D}_\delta^\epsilon \cap \tilde{\mathcal{D}}_2$, ceci pour tout $\epsilon > 0$, c'est-à-dire que :

$$\Delta_\epsilon(u,v) = 0$$

Ainsi :

$$\lim_{\epsilon\to 0} \Delta_\epsilon(u,v) = \mathbf{1}_{(u,v)\in\mathcal{D}_\delta\cap\tilde{\mathcal{D}}_2}$$

Étape 3 : montrons que $\dfrac{I_\epsilon - I}{\epsilon}$ a une limite quand ϵ tend vers zéro et calculons sa limite. Soit K la fonction définie par :

$$K(t) = \int_{l_0}^{l_1}\int_{l_1-x}^{+\infty} \mathbf{1}_{y<t} h(x,y)\mathrm{d}y\mathrm{d}x$$

On a : $K\left(G_\epsilon^{-1}(1-\beta)\right) = I_\epsilon$ et $K\left(G^{-1}(1-\beta)\right) = I$. Montrons que K est dérivable sur $[l_1-l_0;+\infty[$. On se restreint à cet intervalle car on a : $G^{-1}(1-\beta) > l_1 - l_0$, et donc :

$$\exists E > 0, \forall \epsilon < E, G_\epsilon^{-1}(1-\beta) > l_1 - l_0$$

Pour $t \in [l_1 - l_0; +\infty[$:

$$K(t) = \int_{l_0}^{l_1} \left[G_x(t) - G_x(l_1 - x)\right] f(x)\mathrm{d}x$$

Or, pour tout x dans $[l_0; l_1[$, $t \to G_x(t)$ est dérivable sur $[-x;+\infty[$ donc sur $[l_1 - l_0; +\infty[$, et de dérivée égale à : $\dfrac{\partial}{\partial t}G_x(t) = g(t|x)$, pour $t \in [-x, +\infty[$ et nulle ailleurs. De plus, pour tout t de $[l_1 - l_0; +\infty[$, et pour tout x de $[l_0; l_1[$:

$$g(t|x)f(x) \leq Cf(x)$$

où C est une constante indépendante de t. Donc, par le théorème de dérivabilité sous le signe somme, K est dérivable sur $[l_1 - l_0; +\infty[$, de dérivée égale à :

$$\begin{aligned} K'(t) &= \int_{l_0}^{l_1} g(t|x)f(x)\mathrm{d}x \\ &= g(t) \end{aligned}$$

De plus, d'après le lemme 2, la fonction $\phi_v : \epsilon \to G_\epsilon^{-1}(1-\beta_\epsilon)$ (où $G_\epsilon(t) = (1-\epsilon)G(t) + \epsilon \mathbf{1}_{t\geq v}$) est dérivable en $\epsilon = 0$, et sa dérivée en 0 vaut, pour $v \neq G^{-1}(1-\beta)$:

$$\begin{aligned}\phi_v^{'}(0) &= \frac{(1-\beta) - \beta^{'}(0) - \mathbf{1}_{G^{-1}(1-\beta)\geq v}}{g\left(G^{-1}(1-\beta)\right)} \\ &= \frac{(1-\beta) - \beta^{'}(0) - \mathbf{1}_{(u,v)\in\mathcal{D}_\delta}}{g\left(G^{-1}(1-\beta)\right)}\end{aligned}$$

où $\beta^{'}$ est la dérivée de la fonction $\epsilon \mapsto \beta_\epsilon$. En dérivant l'équation (3.28) par rapport à ϵ on en déduit que $\beta^{'}$ est une fonction constante qui vaut :

$$\beta^{'}(\epsilon) = (1-\beta) - (\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} - \alpha \mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}})$$

Ainsi :

$$\phi_v^{'}(0) = \frac{\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} - \alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} - \mathbf{1}_{(u,v)\in\mathcal{D}_\delta}}{g\left(G^{-1}(1-\beta)\right)}$$

Pour $v = G^{-1}(1-\beta)$, comme $1-\beta_\epsilon$ vérifie l'inégalité (3.29), $\phi_v^{'}(0) = 0$. On en déduit la limite de $\Lambda_\epsilon = \dfrac{K\left(G_\epsilon^{-1}(1-\beta)\right) - K\left(G^{-1}(1-\beta)\right)}{\epsilon}$ quand ϵ tend vers zéro :

$$\begin{aligned}\lim_{\epsilon\to 0}\Lambda_\epsilon &= K^{'}\left(G^{-1}(1-\beta)\right)\phi_v^{'}(0) \\ &= \begin{cases} \mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} - \alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} - \mathbf{1}_{(u,v)\in\mathcal{D}_\delta} & \text{pour } v \neq G^{-1}(1-\beta) \\ 0 & \text{pour } v = G^{-1}(1-\beta) \end{cases}\end{aligned}$$

ce qui s'écrit aussi :

$$\lim_{\epsilon\to 0}\frac{I_\epsilon - I}{\epsilon} = \left[\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} - \alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} - \mathbf{1}_{(u,v)\in\mathcal{D}_\delta}\right]\mathbf{1}_{v\neq G^{-1}(1-\beta)}$$

Finalement, on obtient :

$$\begin{aligned}IC_{T_1,H}((u,v)) &= \frac{1}{(1-\alpha)(1-\alpha_{l_2\,l_0})}\Big\{\mathbf{1}_{(u,v)\in\mathcal{D}_\delta\cap\tilde{\mathcal{D}}_2} \\ &\quad + \left[\mathbf{1}_{(u,v)\in\mathcal{D}_{l_2}} - \alpha\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} - \mathbf{1}_{(u,v)\in\mathcal{D}_\delta}\right]\mathbf{1}_{v\neq G^{-1}(1-\beta)}\Big\} \\ &\quad - \frac{\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}}{1-\alpha_{l_2\,l_0}}\frac{1}{(1-\alpha)(1-\alpha_{l_2\,l_0})}\iint_{\mathcal{D}_\delta\bigcap\mathcal{D}_2} h(x,y)\mathrm{d}x\mathrm{d}y\end{aligned}$$

La relation (3.24) donne l'expression de $IC_{T_1,H}((u,v))$ donnée dans la proposition. ∎

Enfin, la fonction d'influence de T_2 au point H dans la direction (u,v), est donnée par la proposition suivante :

Proposition 4 *La fonction d'influence de l'estimateur tronqué généré par la fonctionnelle T_2, au point de la distribution du vecteur taille individuel, H, dans la direction du point (u,v) dans $\mathcal{D}$, $IC_{T_2,H}$, est égale à :*

$$\begin{aligned} IC_{T_2,H}((u,v)) &= \frac{1}{(1-\alpha')(1-\alpha_{l_2\,l_0})}\left[\mathbf{1}_{(u,v)\in\mathcal{D}_{\delta'}\cap\tilde{\mathcal{D}}_2} + A(u,v)\right] \\ &\quad -\frac{T(H)-\alpha_{l_2}-\beta'(1-\alpha_{l_2\,l_0})}{(1-\alpha')(1-\alpha_{l_2\,l_0})^2}\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} \end{aligned}$$

avec :

$$\begin{aligned} A(u,v) &= -\beta_1 C_1 \mathbf{1}_{l_0+\delta'\neq u} + C_1\mathbf{1}_{l_0+\delta'>u} + (1-\beta_2)C_2\mathbf{1}_{l_1-\delta'\neq u} - C_2\mathbf{1}_{l_1-\delta'>u} \\ &\quad -\beta_3' C_1'\mathbf{1}_{l_1+\delta'\neq u+v} + C_1'\mathbf{1}_{l_1<u+v<l_1+\delta'} - \beta_4' C_2'\mathbf{1}_{l_2-\delta'\neq u+v} + C_2'\mathbf{1}_{l_2-\delta'<u+v<l_2} \end{aligned}$$

où C_1, C_2, C_1', C_2' des constantes positives, indépendantes de u et v.

Preuve : La démarche est la même que pour la démonstration de la proposition précedente.

Étape 1 : on donne l'expression de la fonction d'influence de T_2. Pour cela, on exprime l'ensemble $\mathcal{D}_{\delta'}$ en fonction des lois des tailles et des accroissements en taille. La troncature consiste à prendre, les couples $(X, \Delta X)$ tels que :

$$F(X) \in]\beta_1; 1-\beta_2[$$

avec :

$$\begin{cases} F(l_0+\delta') &= \beta_1 \\ F(l_1-\delta') &= \beta_2 \end{cases} \tag{3.30}$$

Puis, parmi les couples restants, on élimine des couples de la façon suivante : si $\tilde{G}$ est la loi de $\Delta X + X | X \in C$, de densité $\tilde{g}$, et définie sur $[0; +\infty[$ par :

$$\tilde{G}(t) = \int_{l_0}^{l_1} G_x(t-x)f(x)\mathrm{d}x$$

et si on pose :

$$a = \tilde{G}(l_1)$$

on élimine les couples $(X, \Delta X)$ tels que :

$$\tilde{G}(X+\Delta X) \in [0; \alpha_{l_0}+\beta_1'] \cup [a-\beta_2'; a+\beta_3'] \cup [1-\alpha_{l_2}-\beta_4'; 1]$$

avec :

$$\begin{cases} \tilde{G}(l_0+\delta^{'}) &= \alpha_{l_0}+\beta_1^{'} \\ \tilde{G}(l_1-\delta^{'}) &= a-\beta_2^{'} \\ \tilde{G}(l_1+\delta^{'}) &= a+\beta_3^{'} \\ \tilde{G}(l_2-\delta^{'}) &= 1-\alpha_{l_2}-\beta_4^{'} \end{cases} \tag{3.31}$$

$\mathcal{D}_{\delta'}$ est ainsi défini par :

$$\mathcal{D}_{\delta'} = \{(x,y) \in \mathcal{D}; x \in I_F, x+y \in I_{\tilde{G}}\}$$

où :

$$\begin{aligned} I_F &=]F^{-1}(\beta_1); F^{-1}(1-\beta_2)[\\ I_{\tilde{G}} &=]\tilde{G}^{-1}(\alpha_{l_0}+\beta_1^{'}); \tilde{G}^{-1}(a-\beta_2^{'})[\cup]\tilde{G}^{-1}(a+\beta_3^{'}); \tilde{G}^{-1}(1-\alpha_{l_2}-\beta_4^{'})[\end{aligned}$$

De plus :

$$1-\alpha^{'} = (1-\beta_1-\beta_2)(1-\beta_1^{'}-\beta_2^{'}-\beta_3^{'}-\beta_4^{'})$$

La fonction d'influence de T_2 en H dans la direction (u,v) s'écrit :

$$IC_{T_2,H}((u,v)) = \lim_{\epsilon \to 0} \frac{1}{\epsilon(1-\alpha)} \left[\frac{\iint_{\mathcal{D}_{\delta'}^{\epsilon} \cap \tilde{\mathcal{D}}_2} (1-\epsilon)h(x,y)\mathrm{d}x\mathrm{d}y + \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{\delta'}^{\epsilon} \cap \tilde{\mathcal{D}}_2}}{\iint_{\tilde{\mathcal{D}}} (1-\epsilon)h(x,y)\mathrm{d}x\mathrm{d}y + \epsilon \mathbf{1}_{(u,v) \in \tilde{\mathcal{D}}}} - \frac{\iint_{\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y}{(1-\alpha_{l_2\, l_0})} \right]$$

avec :

$$\mathcal{D}_{\delta'}^{\epsilon} = \{(x,y) \in \mathcal{D}; x \in I_{F\,\epsilon}, x+y \in I_{G\,\epsilon}\}$$

où :

$$\begin{aligned} I_{F\,\epsilon} &=]F_{\epsilon}^{-1}(\beta_1); F_{\epsilon}^{-1}(1-\beta_2)[\\ I_{G\,\epsilon} &=]\tilde{G}_{\epsilon}^{-1}(\alpha_{l_0\,\epsilon}+\beta_1^{'}); \tilde{G}_{\epsilon}^{-1}(a_{\epsilon}-\beta_2^{'})[\cup]\tilde{G}_{\epsilon}^{-1}(a_{\epsilon}+\beta_3^{'}); \tilde{G}_{\epsilon}^{-1}(1-\alpha_{l_2\,\epsilon}-\beta_4^{'})[\end{aligned}$$

et :

$$\begin{aligned} F_{\epsilon}(s) &= (1-\epsilon)F(s) + \epsilon \mathbf{1}_{s \geq u} \\ \tilde{G}_{\epsilon}(t) &= (1-\epsilon)\tilde{G}(t) + \epsilon \mathbf{1}_{t \geq u+v} \\ a_{\epsilon} &= \tilde{G}_{\epsilon}(l_1) \\ 1-\alpha_{l_0\,\epsilon} &= (1-\epsilon)(1-\alpha_{l_0}) + \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{l_0}} \\ 1-\alpha_{l_2\,\epsilon} &= (1-\epsilon)(1-\alpha_{l_2}) + \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{l_2}} \end{aligned}$$

On remarque que $\alpha_{l_0\,\epsilon}+\beta_1^{'}$, $a_\epsilon-\beta_2^{'}$, $a_\epsilon+\beta_3^{'}$ et $1-\alpha_{l_2\,\epsilon}-\beta_4^{'}$ vérifient la relation (3.10) du lemme 2.

On note de plus :

$$\begin{aligned}
\tilde{\Delta}(u,v) &= \mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}} \\
\Delta_\epsilon^{'}(u,v) &= \mathbf{1}_{(u,v)\in\mathcal{D}_{\delta'}^{\epsilon}\cap\tilde{\mathcal{D}}_2} \\
I^{'} &= \iint_{\mathcal{D}_{\delta'}\cap\tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y \\
I_\epsilon^{'} &= \iint_{\mathcal{D}_{\delta'}^{\epsilon}\cap\tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

Comme précédemment, on va montrer, que $\Delta_\epsilon^{'}(u,v)$ et que $\dfrac{I_\epsilon^{'}-I^{'}}{\epsilon}$ ont une limite quand ϵ tend vers zéro. Ainsi :

$$IC_{T_2,H}((u,v)) = \frac{1}{(1-\alpha^{'})(1-\alpha_{l_2\,l_0})}\left[\lim_{\epsilon\to 0}\Delta_\epsilon^{'}(u,v)+\lim_{\epsilon\to 0}\frac{I_\epsilon^{'}-I^{'}}{\epsilon}-\frac{I}{1-\alpha_{l_2\,l_0}}\tilde{\Delta}(u,v)\right]$$

Étape 2 : on montre, comme précédemment, que :

$$\lim_{\epsilon\to 0}\Delta_\epsilon^{'}(u,v) = \mathbf{1}_{(u,v)\in\mathcal{D}_{\delta'}\cap\tilde{\mathcal{D}}_2}$$

Démontrons maintenant que $\dfrac{I_\epsilon^{'}-I^{'}}{\epsilon}$ a une limite quand ϵ tend vers zéro et calculons sa limite. On pose, pour cela :

$$L(s_1,s_2,t_1,t_2) = \int_{l_0}^{l_1}\int_{l_1}^{l_2}\mathbf{1}_{x\in]s_1;s_2[}\mathbf{1}_{z\in]t_1;t_2[}h(x,z-x)\mathrm{d}x\mathrm{d}z$$

On suppose que :

$$l_0\le s_1,s_2<l_1,\quad\text{et}\quad l_1\le t_1,t_2<l_2 \tag{3.32}$$

Ainsi :

$$L(s_1,s_2,t_1,t_2) = \int_{s_1}^{s_2}\int_{t_1}^{t_2}h(x,z-x)\mathrm{d}x\mathrm{d}z$$

Or :

$$\begin{aligned}
L(s_1,s_2,t_1,t_2,t_3) &= \int_{l_0}^{s_2}\int_{l_1}^{t_2}h(x,z-x)\mathrm{d}x\mathrm{d}z-\int_{l_0}^{s_2}\int_{l_1}^{t_1}h(x,z-x)\mathrm{d}x\mathrm{d}z \\
&\quad-\int_{l_0}^{s_1}\int_{l_1}^{t_2}h(x,z-x)\mathrm{d}x\mathrm{d}z+\int_{l_0}^{s_1}\int_{l_1}^{t_1}h(x,z-x)\mathrm{d}x\mathrm{d}z
\end{aligned}$$

Or, $(x,z) \mapsto h_i(x, z-x)$ est dérivable sur le domaine $\Omega = \{(x,z); x \in [l_0; l_1[, z \in [-x; +\infty[\}$, donc L est dérivable en tout point (s_1, s_2, t_1, t_2) vérifiant les conditions (3.32).
La dérivée partielle de L par rapport à la première variable en $r = (s_1, s_2, t_1, t_2)$ vaut :

$$\begin{aligned}
\frac{\partial L}{\partial s_1}(r) &= -\int_{l_1}^{t_2} h(s_1, z-s_1)\mathrm{d}z + \int_{l_1}^{t_1} h(s_1, z-s_1)\mathrm{d}z \\
&= -\int_{t_1}^{t_2} h(s_1, z-s_1)\mathrm{d}z \\
&= -\int_{t_1}^{t_2} g(z-s_1|s_1)\mathrm{d}z f(s_1) \\
&= -\left[G_{s_1}(t_2-s_1) - G_{s_1}(t_1-s_1)\right] f(s_1)
\end{aligned}$$

On calcule de même les autres dérivées partielles de L :

$$\begin{cases}
\dfrac{\partial L}{\partial s_2}(r) = \left[G_{s_2}(t_2-s_2) - G_{s_2}(t_1-s_2)\right] f(s_2) \\
\dfrac{\partial L}{\partial t_1}(r) = -\displaystyle\int_{s_1}^{s_2} g(t_1-x|x) f(x)\mathrm{d}x \\
\dfrac{\partial L}{\partial t_2}(r) = \displaystyle\int_{s_1}^{s_2} g(t_2-x|x) f(x)\mathrm{d}x
\end{cases}$$

On obtient donc la différentielle de L, à l'aide de l'expression :

$$dL_r\left(h_1, h_2, h_3, h_4\right) = \frac{\partial L}{\partial s_1}(r)h_1 + \frac{\partial L}{\partial s_2}(r)h_2 + \frac{\partial L}{\partial t_1}(r)h_3 + \frac{\partial L}{\partial t_2}(r)h_4$$

On pose de plus :

$$\varphi(\epsilon) = \left(F_\epsilon^{-1}(\beta_1), F_\epsilon^{-1}(1-\beta_2), \tilde{G}_\epsilon^{-1}(a_\epsilon + \beta_3^{'}), \tilde{G}_\epsilon^{-1}(1-\alpha_{l_2\,\epsilon} - \beta_4^{'}\right)$$

On a :

$$\varphi(0) = (l_0+\delta, l_1-\delta, l_1+\delta, l_2-\delta)$$

De plus, d'après l'équation (3.32) et les lemma 1 et 2, on obtient pour la loi F :

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\left[F_\epsilon^{-1}(\beta_1)\right]_{\epsilon=0} &= \frac{\beta_1 - \mathbf{1}_{F^{-1}(\beta_1)\geq u}}{f\left(F^{-1}(\beta_1)\right)}\mathbf{1}_{F^{-1}(\beta_1)\neq u} \\
\frac{\partial}{\partial\epsilon}\left[F_\epsilon^{-1}(1-\beta_2)\right]_{\epsilon=0} &= \frac{1-\beta_2 - \mathbf{1}_{F^{-1}(\beta_1)\geq u}}{f\left(F^{-1}(1-\beta_2)\right)}\mathbf{1}_{F^{-1}(1-\beta_2)\neq u}
\end{aligned}$$

et pour la loi G :

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\left[\tilde{G}_\epsilon^{-1}(a_\epsilon+\beta_3^{'})\right]_{\epsilon=0} &= \Big[(a+\beta_3^{'})+(\mathbf{1}_{l_1\geq u+v}-a) \\
&\quad -\mathbf{1}_{\tilde{G}^{-1}(a+\beta_3^{'})\geq u+v}\Big]\frac{\mathbf{1}_{\tilde{G}^{-1}(a+\beta_3^{'})\neq u+v}}{\tilde{g}\left(\tilde{G}^{-1}(a+\beta_3^{'})\right)} \\
\frac{\partial}{\partial\epsilon}\left[\tilde{G}_\epsilon^{-1}(1-\alpha_{l_2\,\epsilon}-\beta_4^{'})\right]_{\epsilon=0} &= \Big[(1-\alpha_{l_2}-\beta_4^{'})+(\mathbf{1}_{u+v\in\mathcal{D}_{l_2}}-(1-\alpha_{l_2})) \\
&\quad -\mathbf{1}_{\tilde{G}^{-1}(1-\alpha_{l_2}-\beta_4^{'})\geq u+v}\Big]\frac{\mathbf{1}_{\tilde{G}^{-1}(1-\alpha_{l_2}-\beta_4^{'})\neq u+v}}{\tilde{g}\left(\tilde{G}^{-1}(1-\alpha_{l_2}-\beta_4^{'})\right)}
\end{aligned}$$

On en déduit, avec les équations (3.30) et (3.31), la dérivée de φ en 0 :

$$\begin{aligned}
\varphi^{'}(0) &= \Bigg(\frac{\beta_1-\mathbf{1}_{l_0+\delta'\geq u}}{f\,(l_0+\delta')}\mathbf{1}_{l_0+\delta'\neq u},\frac{1-\beta_2-\mathbf{1}_{l_1-\delta'\geq u}}{f\,(l_1-\delta')}\mathbf{1}_{l_1-\delta'\neq u}, \\
&\quad \frac{\beta_3^{'}-\mathbf{1}_{l_1<u+v\leq l_1+\delta'}}{\tilde{g}\,(l_1+\delta')}\mathbf{1}_{l_1+\delta'\neq u+v},\frac{-\beta_4^{'}+\mathbf{1}_{l_2-\delta'<u+v<l_2}}{\tilde{g}\,(l_2-\delta')}\mathbf{1}_{l_2-\delta'\neq u+v}\Bigg)
\end{aligned}$$

Comme : $(L\circ\varphi)^{'}(0)=dL_{\varphi(0)}\left(\varphi^{'}(0)\right)$, on en déduit $(L\circ\varphi)^{'}(0)$:

$$\begin{aligned}
(L\circ\varphi)^{'}(0) &= -\left[G_{l_0+\delta'}(l_2-l_0-2\delta^{'})-G_{l_0+\delta'}(l_1-l_0)\right](\beta_1-\mathbf{1}_{l_0+\delta'\geq u})\mathbf{1}_{l_0+\delta'\neq u} \\
&\quad \left[G_{l_1-\delta'}(l_2-l_1)-G_{l_1-\delta'}(2\delta^{'})\right](1-\beta_2-\mathbf{1}_{l_1-\delta'\geq u})\mathbf{1}_{l_1-\delta'\neq u} \\
&\quad -\int_{l_0+\delta'}^{l_1-\delta^{'}} g(l_1+\delta^{'}-x|x)f(x)\mathrm{d}x\frac{\beta_3^{'}-\mathbf{1}_{l_1<u+v\leq l_1+\delta'}}{\tilde{g}\,(l_1+\delta')}\mathbf{1}_{l_1+\delta'\neq u+v} \\
&\quad +\int_{l_0+\delta'}^{l_1-\delta^{'}} g(l_2-\delta^{'}-x|x)f(x)\mathrm{d}x\frac{-\beta_4^{'}+\mathbf{1}_{l_2-\delta'<u+v<l_2}}{\tilde{g}\,(l_2-\delta')}\mathbf{1}_{l_2-\delta'\neq u+v}
\end{aligned}$$

Ainsi :

$$\begin{aligned}
(L\circ\varphi)^{'}(0) &= -\beta_1C_1\mathbf{1}_{l_0+\delta'\neq u}+C_1\mathbf{1}_{l_0+\delta'>u}+(1-\beta_2)C_2\mathbf{1}_{l_1-\delta'\neq u}-C_2\mathbf{1}_{l_1-\delta'>u} \\
&\quad -\beta_3^{'}C_1^{'}\mathbf{1}_{l_1+\delta'\neq u+v}+C_1^{'}\mathbf{1}_{l_1<u+v<l_1+\delta'}-\beta_4^{'}C_2^{'}\mathbf{1}_{l_2-\delta'\neq u+v}+C_2^{'}\mathbf{1}_{l_2-\delta'<u+v<l_2}
\end{aligned}$$

avec $C_1, C_2, C_1^{'}, C_2^{'}$ des constantes positives, indépendantes de u et v. Comme :

$$\lim_{\epsilon\longrightarrow 0}\frac{I_\epsilon^{'}-I^{'}}{\epsilon}=(L\circ\varphi)^{'}(0)$$

la fonction d'influence au point (u, v) s'écrit :

$$\begin{aligned} IC_{T_2,H}((u,v)) &= \frac{1}{(1-\alpha')(1-\alpha_{l_2\,l_0})}\left[\mathbf{1}_{(u,v)\in\mathcal{D}_{\delta'}\cap\tilde{\mathcal{D}}_2} + (L\circ\varphi)'(0)\right] \\ &\quad -\frac{\mathbf{1}_{(u,v)\in\tilde{\mathcal{D}}}}{1-\alpha_{l_2\,l_0}}\frac{1}{(1-\alpha')(1-\alpha_{l_2\,l_0})}\iint_{\mathcal{D}_{\delta'}\cap\tilde{\mathcal{D}}_2} h(x,y)\mathrm{d}x\mathrm{d}y \end{aligned}$$

■

Remarque 5 *: si on se place dans le cas où $\alpha' = 0$, $(L\circ\varphi)'(0)$ est nulle et on retrouve l'expression de la fonction d'influence dans le cas non tronqué, soit : $IC_{T_2,H} = IC_{T_{l_2\,l_0},H}$.*

Étude de la sensibilité des paramètres

On remarque que la fonction d'influence pour chaque estimateur est constante, non nulle, sur au moins deux domaines dont l'intersection des adhérences est non vide. La valeur de la fonction d'influence est de plus différente sur ces deux domaines. On en déduit que la sensibilité locale pour tous les estimateurs est infinie. La sensibilité locale ne permet donc pas de comparer la robustesse des estimateurs considérés. On se focalise donc sur la sensibilité globale des estimateurs.

On déduit du calcul de la fonction d'influence et des relations de la remarque 3, les sensibilités globales $\gamma_{M\,M'}$, γ et γ_1 respectivement de $T_{M\,M'}$, T et T_1 :

$$\begin{aligned} \gamma_{MM'} &= \max\left\{\frac{T(H)-\alpha_M}{(1-\alpha_{M'}-\alpha_M)^2}; \frac{1-\alpha_{M'}-T(H)}{(1-\alpha_{M'}-\alpha_M)^2}\right\} \\ \gamma &= \max\{T(H); 1-T(H)\} \\ \gamma_1 &= \max\left\{\frac{T(H)-\alpha_{l_2}}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2}; \frac{1-\alpha_{l_0}-T(H)}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2}\right\} \end{aligned}$$

En effet, on remarque au préalable dans l'expression de la fonction d'influence de T_1 que :

$$\frac{T(H)-\alpha_{l_2}-\alpha(1-\alpha_{l_2\,l_0})}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2} \leq \frac{T(H)-\alpha_{l_2}}{(1-\alpha)(1-\alpha_{l_2\,l_0})^2}$$

La comparaison des sensibilités globales est donnée par la proposition suivante :

Proposition 5

$$\gamma_{MM'} \geq \gamma$$
$$\gamma_1 \geq \gamma_{l_2 l_0}$$

Preuve : La comparaison entre $\gamma_{MM'}$ et γ est donnée dans l'annexe C. D'autre part, comme $1-\alpha \leq 1$, on en déduit que $\gamma_1 \geq \gamma_{l_2\, l_0}$.

D'autre part, la fonction d'influence de T_2 ne s'exprime pas de façon simple : elle est constante sur plusieurs domaines différents, la valeur de la constante étant différente sur chaque domaine. De plus, ces valeurs dépendent de constantes qui sont fonctions de la loi des accroissements. On ne peut donc pas exprimer et comparer de façon simple la sensibilité de cet estimateur.

Variance asymptotique et biais de l'estimateur

Dans cette partie, on compare les variances asymptotiques des estimateurs et leurs espérances. Les variances asymptotiques des estimateurs engendrés par T, $T_{MM'}$, T_1 sont respectivement égales à :

$$\begin{aligned} V &= T(H)(1-T(H)) \\ V_{MM'} &= \frac{(T(H)-\alpha_M)\,(1-\alpha_{M'}-T(H))}{(1-\alpha_{MM'})^3} \\ V_1 &= \frac{(T(H)-\alpha_{l_2})\,(1-\alpha_{l_0}-T(H))}{(1-\alpha)^2(1-\alpha_{l_2\, l_0})^3} \end{aligned}$$

Preuve : Le détail du calcul de V et de $V_{M\,M'}$ est dans l'annexe C. De plus, la variance de T_1 est égale à :

$$\begin{aligned} V_1 &= \iint_{\tilde{\mathcal{D}}_1} \frac{(T(H)-\alpha_{l_2})^2}{(1-\alpha)^2(1-\alpha_{l_2\, l_0})^4} h(x,y)\mathrm{d}x\mathrm{d}y \\ &+ \iint_{(x,y)\in\tilde{\mathcal{D}}_2, y\neq l_2-l_0-\delta} \frac{(1-\alpha_{l_0}-T(H))^2}{(1-\alpha)^2(1-\alpha_{l_2\, l_0})^4} h(x,y)\mathrm{d}x\mathrm{d}y \\ &+ \iint_{(x,y)\in\tilde{\mathcal{D}}_2, y=l_2-l_0-\delta} \frac{(T(H)-\alpha_{l_2}-\alpha(1-\alpha_{l_2\, l_0}))^2}{(1-\alpha)^2(1-\alpha_{l_2\, l_0})^4} h(x,y)\mathrm{d}x\mathrm{d}y \end{aligned}$$

L'ensemble $\left\{(x,y)\in\tilde{\mathcal{D}}_2, y=l_2-l_0-\delta\right\}$ étant de mesure nulle par rapport à la mesure relative à la densité h, on en déduit, d'après l'expression de $V_{l_2\, l_0}$, que :

$$V_1 = \frac{V_{l_2\, l_0}}{(1-\alpha)^2}$$

■

De même que pour la sensibilité, le calcul de la variance de $\hat{q}_2^{\bullet}$ amène à distinguer de nombreux cas particuliers, et s'exprime en fonction de la loi des accroissements.

La différence entre V et $V_{M\,M'}$ montre que son signe dépend de $T(H)$ et $\alpha_{M\,M'}$, mais la condition n'est pas simple. Par contre, la variance de T_1 est plus grande que celle de $T_{l_2\,l_0}$.

D'autre part, l'estimateur de $q_\infty^\bullet$ est non biaisé, alors que les estimateurs tronqués le sont. Néanmoins, ils sont tous asymptotiquement non biaisés. De plus, la comparaison entre les espérances des estimateurs découle des relations entre les fonctionnelles, les estimateurs considérés étant des estimateurs empirique de la moyenne. La différence entre les moyennes des estimateurs de $q_\infty^\bullet$ et de $q_{M\,M'}^\bullet$ sont détaillés dans l'annexe C. Le signe dépend des proportions d'individus éliminés. D'autre part, d'après l'équation (3.26) la différence entre les estimateurs de $q_1^\bullet$ et $q_{l_2\,l_0}^\bullet$ est égale à :

$$\mathbb{E}(\hat{q}_1^\bullet) - \mathbb{E}(\hat{q}_{l_2\,l_0}^\bullet) \;=\; \frac{\alpha}{1-\alpha}\,[T_{l_2\,l_0}(H) - 1]$$

Comme $T_{l_2\,l_0}(H) \leq 1$, l'espérance de $\hat{q}_1^\bullet$ est plus petite que celle de $\hat{q}_{l_2\,l_0}^\bullet$.

3.2.3 Application à Paracou

Dans le cas du traitement des données ne vérifiant pas l'hypothèse de Usher, les résultats sont appliqués à un jeu de données sur le site expérimental de Paracou en Guyane Française. Une analyse exploratoire montre que les distributions du diamètre et de l'accroissement diamétrique des arbres sont bien approchées par des distributions exponentielles. Les expressions précises sont données dans l'annexe C. On cherche ici à relâcher l'hypothèse de Usher en utilisant l'estimateur où toutes les données sont conservées. Cette hypothèse dépend des classes d'état. On compare alors les propriétés de l'estimateur classique et de l'estimateur non tronqué en fonction de la taille des classes, le pas de temps étant fixé à deux ans.

Plus précisément, on compare la robustesse, quantifiée par la sensibilité, et l'erreur quadratique asymptotique normalisée des estimateurs considérés. Les détails et les résultats de la procédure sont donnés dans l'annexe C, les résultats étant donnés dans la figure 2 de cette annexe. Ces résultats montrent que l'estimateur non tronqué est plus robuste et son erreur quadratique est plus faible que pour l'estimateur classique, cela quelle que soit la largeur des classes. Néanmoins, à partir d'une certaine largeur de classe, la différence entre les deux estimateurs n'est plus significative. Cela résulte du fait que plus la largeur des classes augmente, plus la proportion d'individus qui violent l'hypothèse de Usher est petite. A contrario, une largeur de classe trop petite entraîne une proportion trop grande, et rend le modèle peu réaliste.

3.3 Conclusion

Dans ce chapitre, on s'est intéressé aux questions suivantes :

- quelle est la meilleure façon de traiter les données qui violent l'hypothèse de Usher (les éliminer ou les lisser) ?
- la question de l'hypothèse de Usher étant traitée, la troncature des tailles et/ou des accroissements extrêmes permet-elle de minimiser l'impact des erreurs de classification des individus sur leur dynamique ?

Ces questions ont été traitées en comparant la robustesse des différents estimateurs des taux de transition découlant de ces différentes approches. Pour la première question, on a montré que l'estimateur non tronqué, où aucune donnée n'est éliminée est plus robuste que l'estimateur classique, qui est tronqué. Néanmoins, la variance et l'espérance de cet estimateur peuvent être plus grandes que celles de l'estimateur classique. Pour la deuxième question, la troncature sur les valeurs extrêmes des accroissements a donné un estimateur moins robuste que l'estimateur classique. En fait, les erreurs sur les accroissements ont une influence sur la dynamique s'ils impliquent une erreur sur la classification des individus. Or, les estimateurs ne sont pas reliés à la classification des individus. La troncature sur l'échantillon apparaît alors comme une perte d'information.

Dans ce chapitre, on a supposé que le pas de temps et les classes d'état étaient fixés. Or pour des largeurs de classes trop grandes, l'apport de l'estimateur non tronqué est peu significatif et un choix de largeur des classes trop petite rend l'estimateur non tronqué plus robuste. Mais le modèle apparaît alors peu réaliste, puisqu'une grande proportion d'individus est rectifiée. La recherche d'estimateurs robuste est ainsi liée au choix des classes d'état. De façon plus générale, le choix des classes d'état est une question à part entière qui sera discutée dans le chapitre de conclusion.

Chapitre 4

Distribution asymptotique dans le modèle densité-dépendant

Introduction

Ce chapitre est un résumé des résultats de l'article présenté en annexe D. On élargit au modèle densité-dépendant les résultats du chapitre 2 sur le comportement asymptotique des prédictions. On pose l'hypothèse que, les paramètres restant dans un domaine inclus dans l'espace des paramètres, pour chaque valeur des paramètres il existe un unique état stationnaire N_∞. Le but de ce chapitre est de déterminer une ellipsoïde de confiance autour de l'estimation de N_∞. Pour cela, la loi asymptotique de l'estimateur du maximum de vraisemblance de N_∞, lorsque les observations sont indépendantes, est déterminée.

4.1 Modèle densité-dépendant

4.1.1 Modèle mathématique

Les notations non définies ici sont reprises du chapitre 1 (et en particulier la partie 1.1). On se place, dans un souci de simplification, dans le cas d'une fécondité moyenne. Les résultats peuvent être étendus de façon directe au cas d'une fécondité par classe. On suppose que les taux de transition du modèle ont la forme d'une fonction de Ricker, c'est-à-dire exponentielle. Une analyse exploratoire sur les données de Paracou a montré en effet que ce modèle s'ajustait bien. Le modèle le plus général que l'on considère est obtenu lorsque

les taux de transition ont chacun leur propre dépendance, c'est-à-dire :

$$\begin{aligned} p_{it} &= \mu_i \exp(-\xi_i N(t).\mathbf{a}) \quad (i = 1...I) \\ q_{it} &= \nu_i \exp(-\kappa_i N(t).\mathbf{a}) \quad (i = 2...I) \\ f_t &= \alpha \exp(-\beta N(t).\mathbf{a}) \end{aligned} \tag{4.1}$$

où p_{it},q_{it}, et f_t sont les taux de transition du modèle au temps t. Le modèle décrit par les équations 4.1 est appelé le modèle complet. Il est composé de $4I$ paramètres. Un modèle plus parcimonieux est obtenu en identifiant des paramètres. En particulier, des études antérieures (Favrichon, 1998) ont montré qu'un modèle adéquat pour le site de Paracou était obtenu en considérant les paramètres ξ_i tous égaux, les paramètres κ_i tous égaux, et les μ_i and ν_i ayant une dépendance polynômiale de degré deux en i :

$$\begin{aligned} p_{it} &= (a_0 + a_1 i + a_2 i^2) \exp(-\xi N(t).\mathbf{a}) \quad (i = 1...I) \\ q_{it} &= (b_0 + b_1 i + b_2 i^2) \exp(-\kappa N(t).\mathbf{a}) \quad (i = 2...I) \\ f_t &= \alpha \exp(-\beta N(t).\mathbf{a}) \end{aligned} \tag{4.2}$$

Ce dernier modèle (4.2) est appelé le modèle simplifié. il est composé de 10 paramètres.

4.1.2 Prédictions du modèle

Le modèle densité-dépendant conduit à différents types de comportement asymptotique. On supposera ici que les paramètres restent dans un domaine de l'espace des paramètres, et que pour chaque valeur de paramètre dans ce domaine, il y a un unique point d'équilibre. Cette hypothèse est vérifiée empiriquement en traçant les diagrammes de bifurcation du modèle. Sous ces hypothèses, $N(t)$ tend vers N_∞ lorsque le temps tend vers l'infini, qui vérifie l'équation :

$$N_\infty = U(N_\infty) N_\infty$$

Le vecteur stationnaire N_∞ est alors défini comme une fonction implicite des paramètres du modèle. Notons θ le vecteur des paramètres de longueur n_θ. La matrice de Usher U étant une fonction des paramètres, on identifie $U(N_\infty, \theta)$ à $U(N_\infty)$. Ainsi, N_∞ est définie de façon implicite par l'équation suivante :

$$\Psi(N_\infty, \theta) = 0$$

où Ψ est l'application de $\mathbb{R}^I \times \mathbb{R}^{n_\theta}$ dans $\mathbb{R}^I$ définie par :

$$\Psi(N, \theta) = N - U(N)\, N \tag{4.3}$$

4.1.3 Modèle statistique

Dans le modèle densité-dépendant, une observation peut être distinguée suivant la parcelle à laquelle elle appartient, et suivant le temps où cette observation est prise. Pour simplifier, on suppose ici que les données ne sont pas longitudinales. Les observations peuvent appartenir à l'une des K parcelles différentes supposées indépendantes. On note n_k le nombre d'observations dans chaque parcelle tel que $\sum_{k=1}^{K} n_k = n$. La loi des observations peut être vue comme un modèle de mélange, où les proportions du modèle sont les proportions du nombre d'individus dans chaque parcelle. Ces proportions sont supposées ici fixées.

Une observation est donc définie comme un triplet (i, j, k) où k est la parcelle à laquelle l'individu appartient, i est sa classe d'état au temps initial et j sa classe d'état au temps suivant. L'état mort est notée †. L'ensemble des observations est $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, où :

$$\begin{aligned}
\mathcal{A}_1 &= \{(i,i,k), 1 \leq i \leq I, 1 \leq k \leq K\} \\
\mathcal{A}_2 &= \{(i,i+1,k), 1 \leq i \leq I-1, 1 \leq k \leq K\} \\
\mathcal{A}_3 &= \{(i,\dagger,k), 1 \leq i \leq I, 1 \leq k \leq K\} \\
\mathcal{A}_4 &= \{(0,1,k), 1 \leq k \leq K\}
\end{aligned}$$

Soit X un vecteur aléatoire dans $\mathcal{A}$. Il décrit l'état des individus au temps initial et au temps suivant, et la parcelle à laquelle l'individu appartient. Conditionnellement au fait qu'il appartienne à la parcelle k, la loi de X est décrite par :

$$\begin{aligned}
\Pr[X = (i,i,k)] &= (1 - f_k^\star)\, p_{ik}\, d_{ik} \\
\Pr[X = (i,i+1,k)] &= (1 - f_k^\star)\, q_{ik}\, d_{ik} \\
\Pr[X = (i,\dagger,k)] &= (1 - f_k^\star)\,(1 - p_{ik} - q_{ik})\, d_{ik} \\
\Pr[X = (0,1,k)] &= f_k^\star
\end{aligned} \tag{4.4}$$

où d_{ik} est la proportion d'individu de la parcelle k dans l'état i ($\sum_{i=1}^{I} d_{ik} = 1$, $\forall k$), p_{ik} et q_{ik} sont les proportions d'individus de la parcelle k qui respectivement restent dans la classe i et passent dans la classe supérieure, et $f_k^\star$ est la proportion d'individus du sous-échantillon de la parcelle k qui est recrutée. Étant donnée la quantité $N_k.\mathbf{a}$ de la parcelle k, l'expression des paramètres p_{ik} et q_{ik} relatifs à la parcelle k est obtenue en remplaçant $N_t.\mathbf{a}$ par $N_k.\mathbf{a}$ dans les expressions (4.1) et (4.2) de p_{it} et q_{it}. D'autre part, f_k est le taux

moyen de reproduction des individus de la parcelle k. La relation entre f_k et $f_k^\star$ est donnée par l'équation :

$$f_k^\star = \frac{f_k}{1+f_k}$$

Un échantillon de taille $n = \sum_{k=1}^{K} n_k$ est la réunion de K sous-échantillon indépendants, le $k^{\text{ème}}$ sous-échantillon étant de taille n_k, et suivant la loi décrite par les équations (4.4).

4.1.4 Estimation des paramètres

On estime les paramètres par maximum de vraisemblance. Étant donné un échantillon $(X_1, \ldots, X_n)$, la vraisemblance du modèle L_θ vérifie $\ln L_\theta(X_1, ..., X_n) = \sum_{j=1}^{n} \ln \ell_\theta(X_j)$, avec

$$\begin{aligned}\ell_\theta(x) \;=\; \sum_{k=1}^{K} \rho_k \Bigg\{ \sum_{i=1}^{m} &(1-f_k^\star) p_{ik} d_{ik} \mathbf{1}_{x=(i,i,k)} + (1-f_k^\star) q_{ik} d_{ik} \mathbf{1}_{x=(i,i+1,k)} \\ &+ (1-f_k^\star)(1-p_{ik}-q_{ik}) d_{ik} \mathbf{1}_{x=(i,\dagger,k)} + f_k^\star \mathbf{1}_{x=(0,1,k)} \Bigg\}\end{aligned} \tag{4.5}$$

où ρ_k est la proportion fixée d'observations dans la parcelle k, et par convention $q_{Ik} = 0$. L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est le point en lequel L_θ est maximum. En fait, la forme développée de la log-vraisemblance montre que la fécondité moyenne peut-être estimée indépendamment des autres paramètres. De plus, dans le modèle complet, les paramètres de chaque classe peuvent être estimés indépendamment des paramètres relatifs aux autres classes. Dans le modèle simplifié, ce n'est plus possible, tous les états dépendant de paramètres communs (a_l et b_l pour $l = 0, 1, 2$).

La maximisation de la vraisemblance est faite de façon numérique en utilisant l'algorithme de Nelder-Mead (Nelder and Mead, 1965). L'avantage du modèle complet est que la maximisation se fait sur quatre paramètres au maximum : $I - 1$ optimisations se font sur quatre paramètres et deux optimisations sur deux paramètres. Dans le modèle simplifié, la maximisation se fait sur huit paramètres simultanément, et une optimisation sur deux paramètres. Le fait d'être piégé dans un maximum local lors de la procédure de maximisation a plus de chance d'avoir lieu dans le modèle simplifié que dans le modèle complet.

Cependant, le modèle simplifié a des avantages par rapport au modèle complet : si aucune donnée n'est disponible dans une classe, les paramètres de cette classe dans le modèle complet ne peuvent pas être estimés. Dans

le modèle simplifié, les paramètres, comme ils ne dépendent pas des classes, sont bien estimés dans ce cas.

4.1.5 Test de la densité-dépendance et selection de modèle

La densité-dépendance peut être testée en testant si les paramètres ξ_i, κ_i, β du modèle complet, et les paramètres ξ, κ, β du modèle simplifié sont nuls. Ce test se fait en utilisant le test du rapport de vraisemblance. On teste $2I$ hypothèses nulles dans le modèle complet : $\xi_i = 0$ $(1 \leq i \leq I)$, $\kappa_i = 0$ $(1 \leq i \leq I-1)$, et $\beta = 0$. Dans le cas du modèle simplifié, on teste $\xi = 0$, $\kappa = 0$ et $\beta = 0$. De plus, $a_2 = 0$ et $b_2 = 0$ sont aussi testés pour justifier le second ordre polynômial dans l'expression des paramètres du modèle simplifié.

De plus, le critère bayésien (BIC) et le critère d'Akaike (AIC) ont été utilisés pour comparer le modèle simplifié et le modèle complet, dans le but de choisir le plus informatif.

4.2 Loi asymptotique des prédictions du modèle

La prédiction du modèle considérée ici est le vecteur stationnaire de distribution en état, N_∞. On suppose qu'il existe et qu'il est une fonction implicite des paramètres $h(\theta)$ définie par l'équation implicite : $\Psi[h(\theta), \theta] = 0$ où Ψ est donnée par l'équation (4.3). Si $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ pour un échantillon de taille n, $\hat{N}_{\infty n} = h(\hat{\theta_n})$ est l'estimateur du maximum de vraisemblance de N_∞. Le comportement asymptotique de $\hat{N}_{\infty n}$ est alors donné par la proposition suivante.

Proposition 6 *Avec les notations ci-dessus,*

- *supposant qu'il existe un voisinage ϑ_θ de θ dans l'espace des paramètres tel que, $\forall \theta' \in \vartheta_\theta$, le modèle de Usher densité-dépendant de paramètres θ' admet un unique point d'équilibre $h(\theta')$,*
- *supposant que, étant donnée la loi des observations, l'ensemble des valeurs possible pour $\hat{\theta}_n$ est inclus dans ϑ_θ,*

alors $\sqrt{n}(\hat{N}_{\infty n} - N_\infty)$ converge en loi vers une loi multinormale dans $\mathbf{R}^I$ de moyenne zéro et de matrice de covariance :

$$\mathbf{S}^2 = (\mathrm{d}_\theta h) \cdot I(\theta)^{-1} \cdot {}^{\mathrm{t}}(\mathrm{d}_\theta h)$$

où ${}^{t}\mathbf{M}$ *est la transposée de la matrice* $\mathbf{M}$*,* $I(\theta)$ *est l'information de Fisher de* θ *pour un échantillon de taille un,* $\mathrm{d}_\theta h$ *est la différentielle de* h *en* θ*, représentée par une matrice de dimension* $I \times n_\theta$*, et le point indique la multiplication matricielle .*

La preuve de cette proposition se fait à l'aide du théorème central limite et de la delta-méthode. Elle s'appuie en particulier sur l'hypothèse d'indépendance des individus et sur l'existence et l'unicité d'un point d'équilibre.

4.3 Application à Paracou

L'étude du modèle densité-dépendant est appliquée à un jeu de donnée sur le site expérimental de Paracou en Guyane Française. Les résultats sont présentés dans l'annexe D. En particulier, l'estimation des paramètres est calculée en maximisant la vraisemblance par des algorithmes d'optimisation. Néanmoins, dans le modèle simplifié, l'optimisation dépend très fortement des paramètres de départ. Il faut donc tout d'abord ajuster ces paramètres de départ pour s'assurer de ne pas être piéger dans des maxima locaux. Le modèle complet ne présente pas ce genre de difficulté car les optimisations se font sur moins de paramètres à la fois (au maximum quatre pour le modèle complet contre huit paramètres pour le modèle simplifié).

4.4 Conclusion

Dans ce chapitre, on a proposé un modèle densité-dépendant où les observations dépendaient uniquement des parcelles dans lesquelles ces observations étaient prises, mais n'étaient pas longitudinales. Les parcelles étant, de plus, supposées indépendantes et les proportions d'individus dans chaque parcelle étant fixées, les variables aléatoires du modèle statistique sont indépendantes. Sous ces hypothèses, l'estimateur du maximum de vraisemblance des paramètres a un comportement asymptotique connu, ainsi que celui de l'estimateur du maximum de vraisemblance du vecteur de distribution stationnaire.

Des difficultés subsistent lorsqu'on veut préciser par des simulations la vitesse de convergence de l'estimateur de N_∞. L'optimisation numérique est en effet très instable, et la procédure d'algorithme reste piégée dans des maxima locaux. Il s'agirait alors de construire des simulations plus performantes ou de re-programmer l'algorithme d'optimisation numérique afin de contourner ces difficultés.

Conclusions et perspectives

Conclusions

L'objectif principal de cette thèse était de construire des intervalles de confiance autour des prédictions du modèle. La variabilité des prédictions due à l'échantillonnage finie est, en effet, souvent négligée. Une fois ces intervalles de confiance déterminés, un second objectif a été de rendre la précision de ces intervalles la plus petite possible. D'un point de vue appliqué, ces intervalles de confiance ont été construits dans le but de servir de guide dans la gestion d'une exploitation forestière. La précision des intervalles de confiance devait, de plus, être reliée à la taille de l'échantillon, afin de déterminer le nombre d'arbres à inventorier pour avoir une précision donnée sur l'estimation des prédictions.

Dans cette thèse, on s'est tout d'abord consacré au modèle linéaire. On a construit des intervalles de confiance asymptotiques des prédictions en utilisant l'estimateur du maximum de vraisemblance. Puis, on a cherché à améliorer cet estimateur par des estimateurs plus robustes. On a alors proposé des estimateurs qui appartenaient à la classe des L-estimateurs. Enfin, les résultats du modèle linéaire, dans le cas de l'estimateur du maximum de vraisemblance, ont été étendus au modèle densité-dépendant.

Dans ce travail, on s'est focalisé sur les prédictions stationnaires du modèle : le taux de croissance et le vecteur de distribution stationnaires. En réalité, l'échelle de temps d'étude est en général d'une dizaine d'années. Mais ces prédictions ont l'avantage d'être facilement calculables. Dans le modèle linéaire, elles correspondent respectivement à la première valeur propre et à son vecteur propre associé. Dans le cas densité-dépendant, parmi l'infinité de matrices de Usher, on se focalise sur celle (supposée existante et unique) dont la valeur propre dominante vaut un. Le vecteur associé est le vecteur stationnaire. En pratique, ces prédictions sont utilisées pour avoir une idée sur la tendance de l'évolution de la population. Néanmoins, d'autres quantités intéressent les forestiers, correspondant aux prédictions du régime transitoire.

Pour construire des intervalles de confiance des prédictions, on a adopté

un point de vue asymptotique en considérant que la population dont on suit l'évolution était de taille infinie, les paramètres étant estimés à partir d'une population de taille finie. Cette approche avait déjà été suivie par Houllier (1986) et Houllier et al. (1989). Les estimateurs que l'on a considérés sont l'estimateur du maximum de vraisemblance et d'autres estimateurs appartenant à la classe des L-estimateurs. Ils ont la propriété commune, les variables aléatoires du modèle statistique considéré étant indépendantes, d'avoir un comportement asymptotique caractérisé par une loi normale centrée, dont la variance est déterminable. Les intervalles de confiance des prédictions obtenus sont alors asymptotiques. Ce travail a permis de préciser par rapport aux résultats apportés par Houllier (1986) et Houllier et al. (1989) que ces résultats étaient asymptotiques et de donner la loi asymptotique des estimateurs.

Des simulations ont montré que les résultats sur les intervalles de confiance étaient bien adaptés à des échantillons de grandes tailles, mais peu précis pour des échantillons de petites tailles. Dans ce dernier cas, l'estimation des paramètres par maximum de vraisemblance peut même s'avérer impossible, si on ne dispose pas de données dans une classe. En foresterie, l'application aux échantillons de petites tailles est pourtant une question importante. La récolte de données est en général très coûteuse et on est souvent face à des données limitées. L'enjeu pour ce type d'échantillon est alors, d'une part, de considérer des estimateurs qui permettent des estimations dans le cas d'absence de données dans une classe (Rogers-Bennett and Rogers, 2006), et, d'autre part, de pouvoir construire des intervalles de confiance qui donnent des résultats satisfaisants pour ce type d'échantillon.

Une autre partie du travail s'est attachée à construire des intervalles de confiance de longueur raisonnable tout en améliorant les estimateurs du point de vue de leur robustesse. Cette approche s'est faite en répondant à deux questions :

- comment traiter les données qui ne respectent pas l'hypothèse de Usher ?
- comment minimiser l'impact des erreurs de mesure sur la dynamique de la population ?

Dans ce travail, on a proposé un estimateur des taux de transition qui consistait à garder toutes les données et on a montré que cet estimateur était plus robuste que l'estimateur classique qui tronquait les données. Néanmoins l'apport de cet estimateur dépend de la largeur des classes. Si la largeur des classes est trop grande, les deux estimateurs ont le même comportement, autant au niveau de leur sensibilité que de leur variance. A contrario, si la largeur des classes est trop petite, la variance des deux estimateurs est trop grande. En tout cas, la recherche d'estimateurs robustes nécessite de préciser au préalable le choix des classes.

Par contre, la recherche d'estimateurs robustes pour limiter l'impact des

erreurs de mesures n'a pas abouti. En fait, les estimateurs proposés sont des estimateurs tronqués de l'estimateur classique. Mais ils ne sont reliés à la valeur des tailles des individus que dans le sens où cette valeur indique l'état de l'individu. Le fait de tronquer l'échantillon représente alors pour ces estimateurs une perte d'information.

Dans le modèle densité-dépendant, la méthode classiquement utilisée est une estimation en deux étapes (Favrichon, 1998). On a présenté une méthode plus rigoureuse de l'estimation des taux de transition en une seule étape, en utilisant l'estimateur du maximum de vraisemblance. Plus encore, cette méthode a apporté en plus des précisions sur la variance de l'estimateur considéré. Celle-ci est approchée par la variance asymptotique. On obtient même la loi asymptotique de l'estimateur. On peut alors en déduire, par la delta-méthode, la loi asymptotique du vecteur de distribution stationnaire.

Néanmoins, le comportement asymptotique des estimateurs des taux de transition a pu être établi parce qu'on a apporté des simplifications au modèle densité-dépendant. Les observations n'étaient pas longitudinales et étaient indépendantes entre elles. On peut se demander comment estimer les taux de transition et quel est le comportement asymptotique de leur estimateur lorsqu'on supprime ces simplifications.

Par ailleurs, l'existence et l'unicité du vecteur de distribution stationnaire a été vérifié par des simulations. Pour être plus rigoureux, il aurait fallu établir de façon analytique le comportement asymptotique du modèle suivant les critères :

- son état asymptotique (état d'équilibre, cycles, chaos) ;
- le nombre d'états d'équilibre éventuels.

Perspectives

De nombreux approfondissements à ce travail restent à faire. Ils sont hiérarchisés suivant qu'ils utilisent les méthodes du travail actuel ou suivant que ces méthodes sont étendues à un cadre plus général.

Le travail actuel peut être étendu dans trois directions au moins :

- on peut utiliser d'autres types d'estimateurs des taux de transition, comme par exemple les estimateurs dans le modèle statistique continu décrit dans le chapitre 1 (Rogers-Bennett and Rogers, 2006) ;
- l'étude des prédictions stationnaires peut être étendu à d'autres types de prédictions du régime transitoire. C'est le cas, par exemple, du temps pour reconstituer le stock à partir d'une petite perturbation (Chagneau, 2006), et des grandeurs agrégées suivantes : l'effectif totale et la surface terrière du régime transitoire. En pratique, le régime transitoire

correspond par exemple à un peuplement après exploitation ;
- on peut faire la même étude dans d'autres modèles matriciels, comme le modèle de Leslie et de Lefkovitch.

Dans tous les cas, le raisonnement reste globalement le même. Pour les estimateurs qui ne sont pas des estimateurs du maximum de vraisemblance, on pourra envisager des estimateurs dont la variance asymptotique est déterminée par la fonction d'influence. La prédiction quand à elle est reliée aux paramètres du modèle, quel que soit le modèle matriciel considéré. La variance asymptotique de la prédiction est une différentielle composée avec la matrice de variance asymptotique des paramètres. Le calcul de la différentielle variera selon la prédiction étudiée.

D'autres perspectives étendent le cadre d'étude actuel. Tout d'abord, le modèle densité-dépendant pourra être élargi au cas où les observations sont longitudinales. Il s'agira de se caler sur un nombre de pas de temps fixé et d'estimer les paramètres sur cet intervalle de temps. On établira alors le comportement asymptotique des estimateurs des paramètres lorsque la taille de la population tend vers l'infini. Ce résultat nécessite au préalable d'établir un résultat analogue au théorème central limite dans ce modèle, dans le cas où les observations ne sont pas indépendantes. D'un point de vue pratique, pour estimer les paramètres du modèle densité-dépendant à partir d'une parcelle suivie dans le temps, on peut se demander si cette façon d'estimer revient à estimer les paramètres de ce modèle à partir de plusieurs parcelles prises au même instant. En allant plus loin, les trajectoires de plusieurs parcelles suivies dans le temps peuvent-elles être mises bout à bout ?

Par ailleurs, les résultats de cette thèse sont des résultats en grande population, où la loi des estimateurs est obtenu lorsque la taille de la population tend vers l'infini. On peut alors s'interroger sur la validité de ces résultats pour des échantillons de petites tailles et comment construire des intervalles de confiance dans ce cas. Dans l'état actuel des travaux, des simulations peuvent être faites pour calculer la variance empirique. On peut aussi envisager d'obtenir des résultats asymptotiques non plus lorsque la taille de l'échantillon tend vers l'infini mais lorsque le nombre de pas de temps sur lesquels les observations sont prises tend vers l'infini. Plus précisément, les paramètres étant estimés sur un nombre de pas de temps fini, il s'agira d'établir le comportement asymptotique des estimateurs lorsque ce nombre tend vers l'infini et conditionnellement au fait que la population ne s'éteint pas en temps fini.

Enfin, dans ce travail, on s'est focalisé sur la stochasticité d'échantillonnage pour estimer les prédictions du modèle. La stochasticité démographique et la stochasticité environnementale sont aussi à prendre en compte. Il s'agirait alors de comparer le poids des différentes stochasticités dans les résultats

obtenus. Dans ce travail, des simulations ont permis de comparer le poids de la stochasticité d'échantillonnage et de la stochasticité démographique dans le calcul de la variance du vecteur de distribution stationnaire. On pourrait aussi étudier plus rigoureusement l'ergodicité de la chaîne de Markov décrivant l'évolution du vecteur d'effectif.

Annexes

A. Article paru dans *Mathematical Biosciences* (avec erratum)
B. Calculs annexes à l'article paru dans *Mathematical Biosciences*
C. Article sous presse dans *Computational Statistics and Data Analysis*
D. Article soumis à *Acta Biotheoretica*

Annexe A

Article paru dans *Mathematical Biosciences*

Erratum

Page 83 de l'article, à la place de :

$$(I(\theta)^{-1})_{i,i} = \frac{p_i(1-p_i)}{d_i} \qquad \text{for } i = 1, \ldots, m$$
$$(I(\theta)^{-1})_{m+i,i} = \frac{p_i q_{i+1}}{d_i} \qquad \text{for } i = 1, \ldots, m-1$$
$$(I(\theta)^{-1})_{m+i,m+i} = \frac{q_{i+1}(1-q_{i+1})}{d_i} \qquad \text{for } i = 1, \ldots, m-1$$
$$(I(\theta)^{-1})_{2m-1+i,2m-1+i} = \frac{f_i}{d_i} \qquad \text{for } i = 1, \ldots, m$$
$$(I(\theta)^{-1})_{3m-1+i,3m-1+i} = \frac{1}{d_i} \qquad \text{for } i = 1, \ldots, m$$

il faut lire :

$$(I(\theta)^{-1})_{i,i} = \frac{p_i(1-p_i)}{d_i} \qquad \text{for } i = 1, \ldots, m$$
$$(I(\theta)^{-1})_{m+i,i} = \frac{-p_i q_{i+1}}{d_i} \qquad \text{for } i = 1, \ldots, m-1$$
$$(I(\theta)^{-1})_{m+i,m+i} = \frac{q_{i+1}(1-q_{i+1})}{d_i} \qquad \text{for } i = 1, \ldots, m-1$$
$$(I(\theta)^{-1})_{2m-1+i,2m-1+i} = \frac{f_i}{d_i} \qquad \text{for } i = 1, \ldots, m$$
$$(I(\theta)^{-1})_{3m-1+i,3m-1+i} = d_i(1-d_i) \qquad \text{for } i = 1, \ldots, m$$
$$(I(\theta)^{-1})_{3m-1+i,3m-1+j} = -d_i d_j \qquad \text{for } i = 1, \ldots, m,\ i \neq j$$

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mathematical Biosciences 200 (2006) 76–89
Mathematical Biosciences
www.elsevier.com/locate/mbs

Asymptotic distribution of stage-grouped population models

M. Zetlaoui [a], N. Picard [b], A. Bar-Hen [a,*]

[a] *Institut National Agronomique Paris-Grignon 16, rue Claude Bernard—75231 Paris Cedex 05, France*
[b] *Cirad-forêt/plantations, Campus international de Baillarguet—TA 10/C 34398, Montpellier Cedex 5, France*

Received 3 May 2005; accepted 5 December 2005
Available online 20 January 2006

Abstract

Matrix models are often used to predict the dynamics of size-structured or age-structured populations. The asymptotic behaviour of such models is defined by their malthusian growth rate λ, and by their stationary distribution w that gives the asymptotic proportion of individuals in each stage. As the coefficients of the transition matrix are estimated from a sample of observations, λ and w can be considered as random variables whose law depends on the distribution of the observations. The goal of this study is to specify the asymptotic law of λ and w when using the maximum likelihood estimators of the coefficients of the transition matrix. We prove that λ and w are asymptotically normal, and the expressions of the asymptotic variance of λ and of the asymptotic covariance matrix of w are given. The convergence speed of λ and w towards their asymptotic law is studied using simulations. The results are applied to a real case study that consists of a Usher model for a tropical rain forest in French Guiana. They permit to assess the number of trees to measure to get a given precision on the estimated asymptotic diameter distribution, which is an important information on tropical forest management.

Keywords: Matrix model; Population dynamics; Asymptotic properties

* Corresponding author.
E-mail addresses: avner@inapg.fr, avner@bar-hen.net (A. Bar-Hen).

0025-5564/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2005.12.003

1. Introduction

Study of dynamics of population is an important tool in many ecological studies. This idea is to propose a simple model that mimic the future evolution of the population. Among the discrete-time models, matrix models are often used to study the dynamics of structured populations (either age-structured or size-structured populations). They also permit to simplify the dynamics of a population into its basic components: recruitment or birth, growth or ageing, and mortality. The Usher model is a matrix model for size-structured populations that is based on four hypothesis [30,31]:

- Hypothesis of independence: the evolution of the individuals are independent.
- Markov hypothesis: the evolution of an individual between two time steps t and $t+1$ only depends on its state at time t.
- Usher hypothesis: during each time step, an individual can stay in the same class, move up a class, or die; each individual may give birth to a number of offspring.
- Hypothesis of stationarity: the evolution of individuals between two time step is independent of time.

More general matrix model was proposed by Lefkovitch [14,20,28] (see also [24] for spatial extension), and allows any transition from one stage to another. The Leslie model [21], which describes a population grouped by age, is a special case of the Usher model: during each time step, an individual can only move up a class or die.

The Usher model is often used in forestry because it permits to simulate quickly and in a synthetic way large areas of forest [10]. This model is particularly adapted to deal with forest management [3], economic potential of forests [4], or biodiversity assessment [5,17,22]. Furthermore, the predictions of the characteristics of the forest stand can be as reliable as the predictions with individual-based models [12,26]. In forestry applications of the Usher model, trees are grouped by diameter class. Diameter is always measured in management-oriented forest inventories.

From a theoretical point of view, many properties of the Usher model are known [6]. In particular its asymptotic behaviour is known: the evolution in the stationary state is exponential and it is characterized by the growth rate λ and the stationary distribution w. The malthusian growth rate λ is often used to diagnostic whether a population will vanish ($\lambda < 1$) or invade ($\lambda > 1$). It is thus important when studying extinction and colonization in population biology. However λ is of less significance in forestry applications, since an undisturbed natural forest is a stationary system ($\lambda = 1$), at least at the scale of a few decades. The most important characteristic for forestry applications is the stationary distribution w that defines the diameter distribution of the forest stand.

Although many theoretical results are available for Usher models, the variability of the predictions due to sampling for parameter estimation has rarely been addressed. It will be proved that it can be a major issue. In this context, the parameters of the Usher model are considered as unknown, and they have to be estimated from a sample of observations [18]. The predicted characteristics such as λ and w are consequently random variables and no longer deterministic variables. Stochasticity due to sampling should not be confused with demographic stochasticity (the evolution of each individual follows a Markov chain) or with environmental stochasticity (the coefficients of the Usher matrix have random fluctuations in time) [32]. In fact, sampling variability

is much more important in terms of variability of the estimation of the parameters. Sampling variability is important for forestry applications, as it relates the quantity of data available for parameter estimation with the precision on the estimation of λ and, more importantly, w. It should permit to answer practical questions such as: How many trees do we have to measure to estimate the asymptotic diameter distribution with a 10% precision at level 5%? Presently, this question cannot be answered in a straightforward manner, but this paper permits to answer it.

The aim of this paper is thus to determine the variance of the estimators of λ and w in Usher models, so as to built confidence interval around the estimations of λ and w. Houllier et al. [18] have determined an approximate expression of the variance and bias of an arbitrary estimator of λ, using a Taylor's development. In this paper, we establish an exact asymptotic result for λ, and, more importantly, for w as well. We focused on the maximum likelihood estimators and we determined their asymptotic law. Then, we obtained the asymptotic variance of the maximum likelihood estimators of λ and w, and we showed that these estimators are asymptotically unbiased. These results readily provide an asymptotic confidence interval for λ and an asymptotic confidence ellipsoid for w. Furthermore, to specify the convergence speed of the estimators of λ and w towards their asymptotic law, we performed simulations. Finally, the results were applied to a data set obtained from experimental forest plots in a tropical rain forest at Paracou (French Guiana).

2. Methods

2.1. The Usher model

The population considered is grouped into m stages. Its evolution is described by the vector $N(t) = (N_i(t))_{i=1,\ldots,m}$, where $N_i(t)$ is the number of individuals in stage i at time t. Time is discrete. Usher's hypothesis says that between time t and $t+1$, an individual either (i) stays in the same stage, or (ii) grows up to the next stage, or (iii) dies. Moving backwards or growing up by two or more stages is not allowed. The deterministic Usher model is [6,30,31]

$$N(t+1) = U(t)N(t).$$

In this paper, U is supposed independent of time. Thus

$$U(t) = U = \begin{pmatrix} p_1 + f_1 & f_2 & \cdots & f_m \\ q_2 & p_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & q_m & p_m \end{pmatrix},$$

where p_i is the probability for an individual to stay in stage i, q_{i+1} is the probability to move up from stage i to $i+1$, and f_i is the fecundity of stage i. The probability of dying for an individual in stage i is $1 - p_i - q_{i+1}$. This implies $p_i \geqslant 0$, $q_{i+1} \geqslant 0$ and $p_i + q_{i+1} \leqslant 1$. We further assume that $p_i > 0$ (otherwise the number m of stages can be reduced), $q_i > 0$ and $f_i > 0$. In some situations it is not possible to estimate a fecundity for each class, but only an average fecundity: $f_1 = \cdots = f_m = f$.

The Usher matrix U is a non-negative primitive matrix. By the Perron–Frobenius theorem, U has a real positive eigenvalue λ of multiplicity equal to one. It is strictly greater in absolute value than all other eigenvalues of U. Let w be the eigenvector associated to λ with a norm equal to one. All the components of w are positive. λ and w determine the asymptotic behaviour of the population: λ is equal to the asymptotic growth rate and w to the stationary diameter structure of the population. Then, if $\lambda < 1$, the population exponentially decreases until extinction; if $\lambda > 1$, it exponentially increases to infinity, and if $\lambda = 1$, it remains in a stationary state. Furthermore, we deduce from the Perron–Frobenius theorem that λ and w are functions of the parameters of U [18]

$$\lambda = h(U) \quad \text{and} \quad w = g(U, h(U)),$$

where h is implicitly defined by

$$\Phi(U, h(U)) = 0 \quad \text{where } \Phi(U, \lambda) = \det(U - \lambda I).$$

g is determined (modulo a multiplicative constant) by the equation

$$Ug(U, \lambda) - \lambda g(U, \lambda) = 0.$$

2.2. Parameter estimation: Maximum likelihood estimator's case

Let $d = (d_1, \ldots, d_m)$ be the state-distribution of the population at initial time t_0, with $\sum_{i=1}^{m} d_i = 1$. Let X be a random vector with values in the set $\mathscr{A} = \{(i, j, l); 1 \leqslant i \leqslant m - 1; j = i, i + 1; l \in \mathbf{N}\} \cup \{(m, m, l); l \in \mathbf{N}\} \cup \{(i, \dagger, l); 1 \leqslant i \leqslant m; l \in \mathbf{N}\}$, where $\dagger$ denotes the death state. The random variable X describes the state of an individual at time t and $t + 1$: (i, i, l) is an individual that stays in class i and gives birth to l offsprings; $(i, i + 1, l)$ is an individual that grows up and gives birth to l individuals; $(i, \dagger, l)$ is an individual that gives birth to l individuals and die. The law F of X is defined as

$$\begin{aligned}
&\Pr[X = (i, i, l)] = p_i d_i g_{il},\\
&\Pr[X = (i, i + 1, l)] = q_{i+1} d_i g_{il},\\
&\Pr[X = (i, \dagger, l)] = (1 - p_i - q_{i+1}) d_i g_{il},
\end{aligned}$$

where g_{il} is the probability that a random variable following a Poisson distribution with parameter f_i be equal to l:

$$g_{il} = \mathrm{e}^{-f_i} \frac{f_i^l}{l!}.$$

To simplify notations, let $\theta = (p_i, 1 \leqslant i \leqslant m; q_i, 2 \leqslant i \leqslant m; f_i, 1 \leqslant i \leqslant m; d_i, 1 \leqslant i \leqslant m) \in \mathbf{R}^{4n-1}$ be the parameters of the law F.

Now, let $(X_1, \ldots, X_n)$ be a sample of size n with values in $\mathscr{A}$, following the distribution $F(\theta)$. In practice, it corresponds to a sample of n individuals observed at two consecutive times, from which the parameters of the Usher matrix U have to be estimated. The model's likelihood is

$$\ell_\theta(x) = \sum_{i=1}^{m} \sum_{l=0}^{\infty} (p_i d_i g_{il} \mathbf{1}_{x=(i,i,l)} + q_{i+1} d_i g_{il} \mathbf{1}_{x=(i,i+1,l)} + (1 - p_i - q_{i+1}) d_i g_{il} \mathbf{1}_{x=(i,\dagger,l)}),$$

where $\mathbf{1}_p$ is the indicator function of proposition p. The maximum likelihood estimator of θ, $\hat{\theta}_n$, is given, for $i = 1, \ldots, m$, by

$$\hat{p}_i = \frac{\sum_{k=1}^{n}\sum_{l=0}^{\infty}\mathbf{1}_{X_k=(i,i,l)}}{N_i},$$

$$\hat{q}_{i+1} = \frac{\sum_{k=1}^{n}\sum_{l=0}^{\infty}\mathbf{1}_{X_k=(i,i+1,l)}}{N_i},$$

$$\hat{f}_i = \frac{\sum_{k=1}^{n}\sum_{l=0}^{\infty} l(\mathbf{1}_{X_k=(i,i,l)} + \mathbf{1}_{X_k=(i,i+1,l)} + \mathbf{1}_{X_k=(i,\dagger,l)})}{N_i},$$

$$\hat{d}_i = \frac{N_i}{n},$$

where

$$N_i = \sum_{k=1}^{n}\sum_{l=0}^{\infty}(\mathbf{1}_{X_k=(i,i,l)} + \mathbf{1}_{X_k=(i,i+1,l)} + \mathbf{1}_{X_k=(i,\dagger,l)}).$$

When an average fecundity f is estimated, rather than a fecundity by class, the maximum likelihood estimator of f simplifies into

$$\hat{f} = \frac{\sum_{k=1}^{n} P(X_k)}{n},$$

where P is the application from $\mathscr{A}$ into $\mathbf{N}$ such that $P[(i,j,l)] = l$. The numerator of $\hat{f}$ is basically the total number of offsprings generated by the n individuals.

2.3. The Paracou data set and simulations

2.3.1. The Paracou data set

The Usher model is used to predict the population dynamics of a forest stand. Data come from the Paracou experimental site in French Guiana (5°15′N, 52°55′W) [13]. It consists of twelve 6.25 ha plots of natural rain forest. The plots were created in 1984. All trees with a diameter at breast height greater than 10 cm, including ingrowth, are monitored. The spatial coordinates and the species of a tree is recorded when it is recruited. Its diameter is measured annually (every two years since 1995), unless it dies. In 1987, nine plots underwent silvicultural treatments with a varying logging intensity. The three remaining plots were left as control. In this study, we use the data from 1984 to 1986 (before logging), or from 1984 to 1997 for control plots only. The data are thus representative of an undisturbed forest.

The time step and the breakpoints of the diameter classes were chosen following Favrichon's [10] recommendations. They insure that Usher's hypothesis is verified. The time step thus is two years. Individuals are grouped into 11 diameter classes ranging from 10 cm to 60 cm, with a constant amplitude equal to 5 cm. The last diameter class groups all the trees with a diameter greater than 60 cm.

The data set consists of 47 301 trees. For each tree, the diameter in 1984 and in 1986 is given, unless the tree was dead in 1986 or not yet recruited in 1984. In the former case, only the diameter in 1984 is given, and in the latter case, only the diameter in 1986 is given. A few individuals (187

trees out of 47 301, that is 0.4%) did not comply with Usher's hypothesis. These data were transformed as follows: trees that moved backwards were considered having stayed in the same class, and trees that grew up by several classes were considered having grown by one class. As the data do not allow to link a recruited tree to its parent, an average fecundity was estimated rather than a fecundity by class.

The maximum likelihood estimates of the parameters of the Usher matrix at Paracou are

$$\begin{cases} (\hat{d}_k)_{k=1\ldots11} = (0.395, 0.211, 0.128, 0.078, 0.054, 0.041, 0.029, 0.019, 0.013, 0.009, 0.022), \\ (\hat{p}_k)_{k=1\ldots11} = (0.949, 0.946, 0.938, 0.934, 0.938, 0.935, 0.909, 0.921, 0.921, 0.930, 0.978), \\ (\hat{q}_k)_{k=2\ldots10} = (0.036, 0.036, 0.044, 0.051, 0.050, 0.050, 0.074, 0.057, 0.062, 0.060), \\ \hat{f} = 0.018. \end{cases}$$

The vector $(\hat{d}_k)_{k=1\ldots11}$ gives the diameter distribution of the population in 1984 and it is shown as black dots in Fig. 1.

2.3.2. Simulations

This paper gives in a theoretical way the asymptotic laws of the estimators of λ and w. To specify the speed of convergence towards the asymptotic law as the sample size increases, we used simulations. We focused on the speed of convergence of two quantities towards their asymptotic value: (i) the variance of the estimator of λ, denoted γ, and (ii) the square root of the first eigenvalue of the covariance matrix of the estimator of w. We denote Γ the covariance matrix of w, and μ the square root of its first eigenvalue. This quantity was used because it is proportional to the length of the greatest principal axis of the confidence ellipsoids [2, p. 59] (the interpretation of the confidence ellipsoids is discussed in the last part). Simulations rely on a law F with known

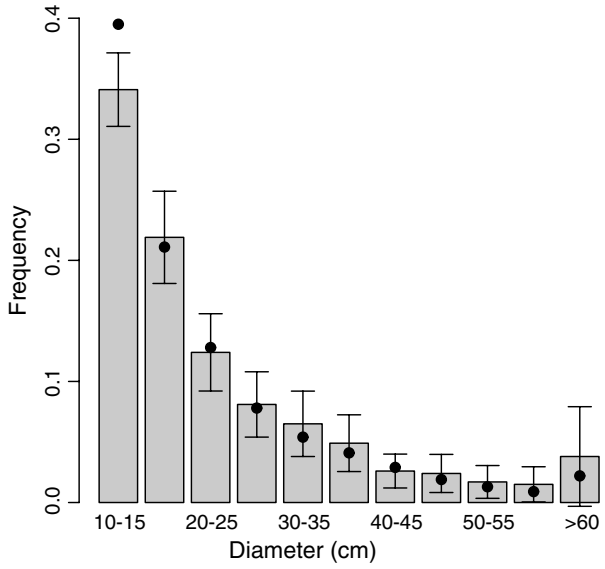


Fig. 1. Diameter distribution at Paracou: bars show the predicted stationary distribution $\hat{w}$, whereas dots show the observed distribution in 1984.

parameters. We actually used as given parameter values for simulations those obtained at Paracou. For a given sample size n, 1000 samples of size n are drawn following the law F. For each sample, an estimate $\hat{\theta}_n$ of the parameters of F is computed. The theoretical asymptotic values of the focus quantities were compared to simulated values to know, on one hand, for which sample size the asymptotic variances of λ and w are well estimated, and on the other hand, for which sample size the asymptotic result applies. Comparisons were made in two ways:

(1) The asymptotic variance of λ, $\gamma(\hat{\theta}_n)$ and the asymptotic covariance matrix of w, $\Gamma(\hat{\theta}_n)$ are computed using the estimated parameters $\hat{\theta}_n$ in place of the true parameters of the law F. For each sample size n, 1000 values of $\gamma(\hat{\theta}_n)$ and 1000 values of $\mu(\hat{\theta}_n)$ are thus obtained. Then we plot the empirical means of these 1000 values against the sample size n.
(2) We compute $\hat{\lambda}_n$ and $\hat{w}_n$ for each sample by diagonalizing the estimated Usher matrix $U(\hat{\theta}_n)$. For each sample size n, 1000 values of $\hat{\lambda}_n$ and 1000 values of $\hat{w}_n$ were thus obtained. We then computed the empirical variance of the 1000 values of $\hat{\lambda}_n$, $\hat{s}_n$, and the empirical covariance matrix of the 1000 values of $\hat{w}_n$, $\hat{S}_n$. Then, we draw the functions $n \mapsto n\hat{s}_n$, and $n \mapsto \sqrt{n}\hat{\mu}_n$, where $\hat{\mu}_n$ is the square root of the first eigenvalue of $\hat{S}_n$. We multiply the value of $\hat{s}_n$ and $\hat{S}_n$ by the sample size because the convergence rate of the maximum likelihood estimator is $\sqrt{n}$.

In simulations, the sample size n varied from 50 to 500 000. Simulations are done using R software [27].

3. Results

3.1. Asymptotic behaviour of λ and w

The dominant eigenvalue of U, λ, is an implicit function of the parameters of U: $\lambda = h(\theta)$. In fact, $\lambda = h[\Pi(\theta)]$ where Π is the projection from $\mathbf{R}^{4m-1}$ to $\mathbf{R}^{3m-1}$ defined as $\Pi : \theta = (p_i, q_i, f_i, d_i)_{1 \leqslant i \leqslant m} \mapsto \Pi(\theta) = (p_i, q_i, f_i)_{1 \leqslant i \leqslant m}$. Then, the maximum likelihood estimator of λ is $\hat{\lambda}_n = h(\Pi(\hat{\theta}_n))$. The asymptotic behaviour of $\hat{\lambda}_n$ is given by the following convergence in law:

Proposition 1. *With the notations above*

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{\mathscr{L}} \mathscr{N}(0, (d_{\Pi(\theta)}h)^t \cdot (d_\theta \Pi)^t \cdot I(\theta)^{-1} \cdot (d_\theta \Pi) \cdot (d_{\Pi(\theta)}h)),$$

where $\xrightarrow{\mathscr{L}}$ denotes the convergence in law, M^t is the transpose of the matrix M, $I(\theta)$ is the Fisher information matrix, $d_a f$ is the differential of f in a, and dot indicates the matrix multiplication.

Proof. The element on the ith row and the jth column of the Fisher information matrix, $I(\theta)$, is defined by

$$(I(\theta))_{ij} = E_\theta \left[\frac{\partial}{\partial \theta_i} (\log \ell_\theta) \frac{\partial}{\partial \theta_j} (\log \ell_\theta) \right].$$

We obtain for $1 \leqslant i \leqslant m$

$$E_\theta\left[\frac{\partial}{\partial p_i}(\log \ell_\theta)\frac{\partial}{\partial p_i}(\log \ell_\theta)\right] = d_i\left(\frac{1}{p_i} + \frac{1}{1-p_i-q_{i+1}}\right),$$
$$E_\theta\left[\frac{\partial}{\partial q_{i+1}}(\log \ell_\theta)\frac{\partial}{\partial q_{i+1}}(\log \ell_\theta)\right] = d_i\left(\frac{1}{q_{i+1}} + \frac{1}{1-p_i-q_{i+1}}\right),$$
$$E_\theta\left[\frac{\partial}{\partial f_i}(\log \ell_\theta)\frac{\partial}{\partial f_i}(\log \ell_\theta)\right] = \frac{d_i}{f_i},$$
$$E_\theta\left[\frac{\partial}{\partial d_i}(\log \ell_\theta)\frac{\partial}{\partial d_i}(\log \ell_\theta)\right] = \frac{1}{d_i},$$
$$E_\theta\left[\frac{\partial}{\partial p_i}(\log \ell_\theta)\frac{\partial}{\partial q_{i+1}}(\log \ell_\theta)\right] = \frac{d_i}{1-p_i-q_{i+1}}.$$

The other elements are equal to zero. $I(\theta)$ is invertible, and its inverse is equal to

$$(I(\theta)^{-1})_{i,i} = \frac{p_i(1-p_i)}{d_i} \quad \text{for } i = 1,\ldots,m,$$
$$(I(\theta)^{-1})_{m+i,i} = \frac{p_i q_{i+1}}{d_i} \quad \text{for } i = 1,\ldots,m-1,$$
$$(I(\theta)^{-1})_{m+i,m+i} = \frac{q_{i+1}(1-q_{i+1})}{d_i} \quad \text{for } i = 1,\ldots,m-1,$$
$$(I(\theta)^{-1})_{2m-1+i,2m-1+i} = \frac{f_i}{d_i} \quad \text{for } i = 1,\ldots,m,$$
$$(I(\theta)^{-1})_{3m-1+i,3m-1+i} = \frac{1}{d_i} \quad \text{for } i = 1,\ldots,m.$$

Consequently the asymptotic behaviour of the maximum likelihood estimator of θ, $\hat{\theta}_n$, is given by

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathscr{L}} \mathscr{N}(0, I(\theta)^{-1}).$$

The implicit function theorem proves that h is differentiable and it gives the expression of its differential. Then the proposition is obtained from the asymptotic behaviour of $\hat{\theta}_n$ using the delta method. □

As before, we obtain the asymptotic behaviour of the maximum likelihood estimator $\hat{w}_n$ of $w = g(U)$:

Proposition 2. *With the same notations as before*

$$\sqrt{n}(\hat{w}_n - w) \xrightarrow{\mathscr{L}} \mathscr{N}(0, (d_{\Pi(\theta)}g)^t \cdot (d_\theta \Pi^t) \cdot I(\theta)^{-1} \cdot (d_\theta \Pi) \cdot (d_{\Pi(\theta)}g)).$$

The proof is the same as before. Houllier et al. [18] had obtained an approximation of the variance of λ with a Taylor's development. We have shown here that this approximation is equal to the asymptotic variance of the maximum likelihood estimator divided by n. Furthermore, we have specified the asymptotic law of the likelihood estimator, that is a centered normal law. In particularly, the bias tends to zero. We have also specified the asymptotic behaviour of the eigenvector

w. Finally, we can deduce that any linear combination of elements of w is asymptotically normal. This is especially interesting for forest management, in particular to calculate cost functions.

3.2. Application to Paracou

The estimate of λ at Paracou is $\hat{\lambda} = 1.002$. The estimate of w at Paracou, $\hat{w}$, is given in Table 2. The graph of $\hat{w}$ is given in Fig. 1.

We remark that $\hat{\lambda}$ is slightly higher that 1. We cannot conclude on the asymptotic behaviour of the population without considering the variance of $\hat{\lambda}$. Furthermore, in view of the graph of w (Fig. 1), we make a χ^2 test to know if the distribution is exponential. The asymptotic distribution is compared to an exponential distribution because the diameter distribution of a natural undisturbed forest is known to be exponential (this result is known in forestry as the de Liocourt law, see [25]). The parameter η of the exponential law is determined as the one that minimizes the test statistic of the χ^2 goodness-of-fit test for an exponential distribution. We find $\eta = 0.072\ \text{cm}^{-1}$, and at level 5%, the test is significant for a sample size higher than 846.

The asymptotic variance of λ and w are estimated using Propositions 1 and 2 respectively. The estimated value of the asymptotic variance of λ is 0.035. The estimated value of the asymptotic covariance matrix of w is given in Table 1. To assess the importance of the variability of w due to sampling, we can compare it to the variability of w due to demographic stochasticity. The latter is obtained when simulating the trajectory of every individual as a Markov chain whose states are the diameter classes [11]. The variability of w due to demographic stochasticity turned to be 100 times smaller than the variability due to sampling. This advocates for the quantification of the variability of the Usher model predictions due to the sampling for parameter estimation.

The asymptotic correlation matrix of w can be computed from its asymptotic variance matrix; its estimate is given in Table 1. The correlation between the first and the other classes is negative. The phenomena is the same for the second class with the others. Thus a perturbation in the first class has an opposite effect on the other classes: an increase in the proportion in the first class brings a decrease in proportion in the other classes, and conversely. On the other hand, the cor-

Table 1
Covariance (upper triangle) and correlation (lower triangle) matrix of $\hat{w}$

	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	$\hat{w}_4$	$\hat{w}_5$	$\hat{w}_6$	$\hat{w}_7$	$\hat{w}_8$	$\hat{w}_9$	$\hat{w}_{10}$	$\hat{w}_{11}$
$\hat{w}_1$	11.35	−0.76	−0.45	−0.36	−0.32	−0.29	−0.26	−0.23	−0.19	−0.16	−0.20
$\hat{w}_2$	−10.87	17.87	−0.04	−0.03	−0.04	−0.03	−0.02	−0.02	−0.02	−0.02	−0.05
$\hat{w}_3$	−5.36	−0.57	12.54	0.20	0.16	0.14	0.13	0.11	0.08	0.06	0.03
$\hat{w}_4$	−3.60	−0.44	2.13	8.96	0.27	0.23	0.22	0.18	0.14	0.10	0.07
$\hat{w}_5$	−3.23	−0.51	1.63	2.36	8.98	0.29	0.27	0.22	0.17	0.13	0.10
$\hat{w}_6$	−2.50	−0.34	1.26	1.79	2.23	6.76	0.36	0.29	0.23	0.19	0.15
$\hat{w}_7$	−1.35	−0.14	0.72	1.00	1.24	1.43	2.41	0.41	0.34	0.27	0.23
$\hat{w}_8$	−1.30	−0.16	0.63	0.89	1.11	1.30	1.10	3.05	0.34	0.27	0.23
$\hat{w}_9$	−0.95	−0.15	0.41	0.59	0.75	0.89	0.77	0.86	2.26	0.35	0.30
$\hat{w}_{10}$	−0.85	−0.17	0.31	0.47	0.61	0.74	0.65	0.73	0.83	2.60	0.33
$\hat{w}_{11}$	−2.88	−1.03	0.35	0.83	1.17	1.60	1.50	1.71	1.98	2.31	20.84

$\hat{w}_i (i = 1, \ldots, m)$ is the ith component of $\hat{w}$.

relation is positive between the other classes: from the third class, a perturbation in a class has the same effect on another class. It is strongest with the second class, and decreases with the class. This decrease may be interpreted in a biological way: as trees grow, they reach the forest canopy and escape from competition for light. Furthermore, we deduce that a perturbation in the first class brings a perturbation in the other classes that decreases from the second class upwards.

The asymptotic behaviour of the estimator $\hat{\lambda}$ provides an asymptotic confidence interval on λ. At level 5%, the confidence interval is: $[1.0019 \pm 0.0017]$. We conclude that 1 is not in the confidence interval. At level 1%, the confidence interval is: $[1.0019 \pm 0.0022]$. For this level, 1 is in the confidence interval. Thus it is difficult to conclude on the stationarity of the population. The value $\hat{\lambda} > 1$ could come from locally (both in space and time) favourable conditions, such as a favourable climate in 1984–1986, a fertile ground or previous logging that would imply that the forest stand is in a growth phase. The forest is in a growth phase in 1984–1986 that could be reverse later on. Actually, if we look at the evolution of the number of trees in the control plots between 1984 and 1997, we observe some fluctuations with an overall decrease of 3% of the number of trees between 1984 and 1997. The Usher model is not suited to model such fluctuations because of the stationarity hypothesis. We have to consider a model with parameters that depend on the characteristics of the population (density-dependent model), time, and environmental variables such as climate variables.

Furthermore, the asymptotic behaviour of $\hat{w}$ gives an asymptotic confidence ellipsoid in $\mathbf{R}^{m-1}$ ($\hat{w} \in \mathbf{R}^m$ but the sum of the components of $\hat{w}$ equal one, so that $\hat{w}$ lies within an hyperplane of $\mathbf{R}^m$). We can readily obtain a confidence interval for each class independently of the other classes using the diagonal of the covariance matrix of $\hat{w}$. The confidence intervals for each class at level 5% are given in Table 2 and are shown in Fig. 1. We remark that the amplitude of the confidence intervals are high, specially for the last classes. Consequently for forest management, it may turn difficult to estimate the number of exploitable trees. Nevertheless these intervals are not sufficient to describe the confidence ellipsoid of w, because they do not consider the dependence between the classes.

Conversely, the asymptotic behaviour of $\hat{w}$ provides the number of trees to measure to reach a given precision in the estimate of w at level 5%. For example, to have a precision at least equal to

Table 2
Estimate and confidence intervals of $\hat{w}$ for each class at Paracou: i is the diameter class, $\hat{w}_i$ the estimate of the ith component of $\hat{w}$, and a_i is the amplitude of the confidence interval for $\hat{w}_i$

i	$\hat{w}_i$	a_i
1	0.341	0.030
2	0.219	0.038
3	0.124	0.032
4	0.081	0.027
5	0.065	0.027
6	0.049	0.023
7	0.026	0.014
8	0.024	0.016
9	0.017	0.013
10	0.015	0.014
11	0.038	0.041

10% for all classes, the sample size must be greater than 8006. Nevertheless, this result is available after studying the convergence speed of the estimator of w.

3.3. Simulation results

With the considered parameters of simulation, noted θ, the asymptotic variance of λ, $\gamma(\theta)$, is: $\gamma(\theta) = 0.036$ and the square root of the first eigenvalue, $\mu(\theta)$, of the asymptotic covariance matrix of w, $\Gamma(\theta)$, is: $\mu(\theta) = 5.312$. The results of simulations are given in Figs. 2 and 3.

Fig. 2 shows the estimations of the asymptotic variance of λ, $\gamma(\hat{\theta}_n)$, and of the square root of the first eigenvalue of the asymptotic covariance matrix of w, $\mu(\hat{\theta}_n)$. The horizontal line indicates the value of $\gamma(\theta)$ (Fig. 2(a)), or of $\mu(\theta)$ (Fig. 2(b)). For a sample size n, we can compare $\gamma(\hat{\theta}_n)$ with $\gamma(\theta)$, and $\mu(\hat{\theta}_n)$ with $\mu(\theta)$. Firstly, the figure confirms the convergence of $\gamma(\hat{\theta}_n)$ towards $\gamma(\theta)$, and of $\mu(\hat{\theta}_n)$ towards $\mu(\theta)$, as n tends to infinity. Furthermore, $\gamma(\hat{\theta}_n)$ is always less than $\gamma(\theta)$ (Fig. 2(a)). It decreases until a sample size of 1000, then grows, and finally reaches a stable value for a sample size greater than 10 000. On the other hand, $\mu(\hat{\theta}_n)$ grows until $n = 1000$, then decreases, and finally stabilizes for a sample size greater than 50 000 (Fig. 2(b)). It is less than $\mu(\theta)$ for $n < 500$.

Fig. 3 shows the empirical variance of λ times n, $n\hat{s}_n$, and the square root of the first eigenvalue of the empirical covariance matrix of w times $\sqrt{n}$, $\sqrt{n}\hat{\mu}_n$. As before, the horizontal line indicates the value of $\gamma(\theta)$ (Fig. 3(a)), or of $\mu(\theta)$ (Fig. 3(b)). In view of this figure, we can compare the value of $n\hat{s}_n$ with $\gamma(\theta)$, and $\sqrt{n}\hat{\mu}_n$ with $\mu(\theta)$. As expected, $n\hat{s}_n$ converges towards $\gamma(\theta)$, and $\sqrt{n}\hat{\mu}_n$ converges towards $\mu(\theta)$. The quantity $n\hat{s}_n$ grows until $n = 5000$, and then oscillates around the value of $\gamma(\theta)$ due to numerical computation inaccuracy (Fig. 3(a)).

In view of Figs. 2 and 3, $\gamma(\theta)/n$ is a good approximation of the variance of $\hat{\lambda}_n$ for sample sizes greater than 5000, but $\gamma(\theta)$ is well estimated for a sample size greater than 10 000. Furthermore, for all sample sizes, $\gamma(\hat{\theta}_n)$ underestimates $\gamma(\theta)$. Similarly, $\mu(\theta)/\sqrt{n}$ is a good approximation of the square root of the first eigenvalue of the covariance matrix of w for sample sizes greater than 5000, but $\mu(\theta)$ is well estimated for a sample size greater than 50 000.

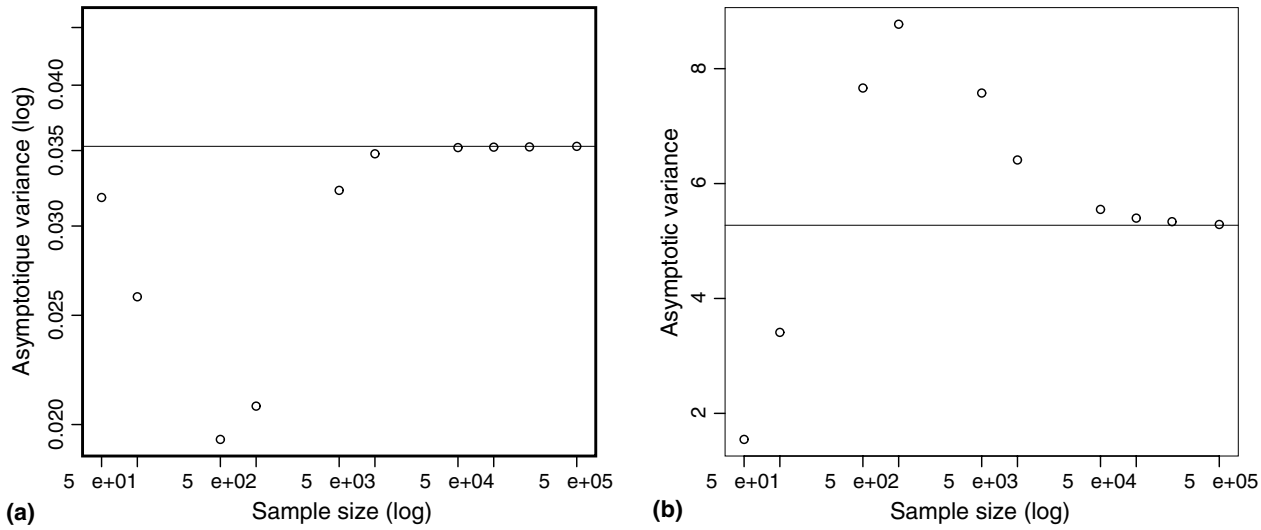


Fig. 2. Estimation of the asymptotic variance of (a) λ and (b) w (1000 simulations).

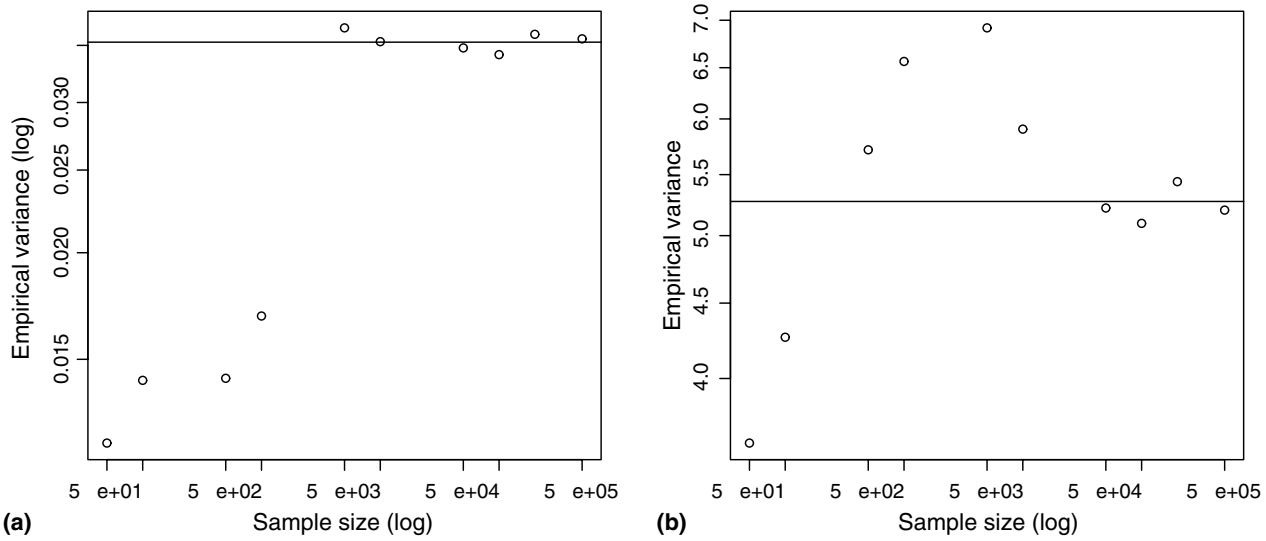


Fig. 3. Empirical variance of (a) λ and (b) w (1000 simulations).

In conclusion, the asymptotic result can be applied to a population of big size such as the forest stand at Paracou, or to abundant species in a large forest stand. However the asymptotic result is not adapted for a small population such as rare species.

4. Discussion

The distribution of the maximum-likelihood estimator of the malthusian growth rate λ of the Usher model is asymptotically normal. Its variance coincides with the expression obtained by Houllier et al. [18] for an arbitrary estimator using a Taylor's expansion. As the maximum-likelihood estimator is asymptotically efficient, it means that the expression given by Houllier et al. may underestimate the asymptotic variance. The application to Paracou data showed that $\hat{\lambda}$ was significantly greater than 1 at the 5% level, which means that the number of individual would infinitely increase. Of course this long-term prediction is unrealistic, as a natural undisturbed forest remains in a stationary state. One could impose a condition on the parameters of the Usher matrix so that $\lambda = 1$. We prefer to use the estimated value of λ as a diagnostic on the forest stand during the period when the parameters were estimated. In the present case, it indicates that growth conditions were favourable at Paracou between 1984 and 1986. The longitudinal data between 1984 and 1997 on control plots confirms this interpretation, as the 1984–1986 period coincided with a growth phase in a globally fluctuating trajectory.

In this paper, we also derived the asymptotic law of the estimator, $\hat{w}_n$, of the stationary distribution w of the population. This quantity is of great interest for foresters as it gives the potentiality of the forest stand for logging. The co-variance matrix of $\hat{w}_n$, $\Gamma(\hat{\theta}_n)$, gives a $(m-1)$-dimensional confidence ellipsoid. We can readily obtain a confidence interval for each class independently of the other classes. To build confidence intervals that consider the dependence between the classes, we can consider the principal axes of the confidence ellipsoid. Their

length are equal to $c\mu_i^{1/2}$, where the μ_is are the eigenvalues of $\Gamma(\hat{\theta}_n)$ and c is the contour level that can be chosen so that the ellipsoid contains 95% of the distribution [2]. We can also consider a paving centered in $\hat{w}$ that contains the ellipsoid, or a paving centered in $\hat{w}$ contained in the ellipsoid.

The quantities λ and $\hat{w}$ are computed from the parameters θ of the Usher matrix. In this paper, we used the maximum-likelihood estimator of θ, and the asymptotic law of the corresponding estimators of λ and w were derived from the asymptotic normal law of θ using the delta method. Nevertheless, other estimators of θ have been proposed, that result in other estimators of λ and w. Michie and Buongiorno [23], for instance, used regression estimators for θ and showed that they were strongly biased compared to the maximum-likelihood estimator. [12] used estimators computed from diameter increments.

The results that we obtained for the maximum-likelihood estimator can be extended to other estimators using the influence function [7,15]. The influence function actually permits to compute the asymptotic variance, not the asymptotic law. Furthermore, the estimator has to be as $T(F_n)$, where F_n is the empirical distribution function of the considered sample and T is a given functional. The maximum-likelihood estimators that we studied in this paper can be written in the form $T(F_n)$, and the asymptotic variance of λ and w that can be computed using the influence function coincides with the expression given before. Another interest of the influence function is that it allows to assess the robustness of the estimators [16,19].

The Usher model assumes that the parameters of the matrix are constant in time. The limits of such an hypothesis are illustrated at Paracou with the value of λ that is significantly greater than 1 at the 5% level. This unrealistic hypothesis has been relaxed using matrix models with density-dependent parameters, or parameters that depend on time [1,8–10,29]. In view of a forest management, these models are particularly interesting to study populations that are subject to perturbations. The estimators of the parameters of the matrix used in density-dependent matrix models are different from the ones used in classical Usher models. Moreover, the asymptotical malthusian growth rate λ is no longer an interesting parameter in density-dependent matrix models [9]. The asymptotic diameter distribution w is the major prediction of the model, and it would be interesting to know its variability due to sampling. We thus intend to extend the results presented in this paper to density-dependent matrix models.

References

[1] L.J.S. Allen, A density-dependent Leslie matrix model, Math. Biosci. 95 (2) (1989) 179.
[2] M. Bilodeau, D. Brenner, Theory of Multivariate Statistics, Series: Springer Texts in Statistics, Hardcover, 1999.
[3] J. Buongiorno, B.R. Michie, A matrix model of uneven-aged forest management, Forest Sci. 26 (4) (1980) 609.
[4] J. Buongiorno, S. Dahir, H.C. Lu, C.R. Lin, Tree size diversity and economic returns in uneven-aged forest stands, Forest Sci. 40 (1) (1994) 83.
[5] J. Buongiorno, J.L. Peyron, F. Houllier, M. Bruciamacchie, Growth and management of mixed-species, uneven-aged forests in the French Jura: Implications for economic returns and tree diversity, Forest Sci. 41 (3) (1995) 397.
[6] H. Caswell, Matrix population models: Construction, Analysis, and Interpretation, second ed., Sinauer Associates, MA, 2001.
[7] A. Cuevas, J. Romo, On the estimation of the influence curve, Can. J. Statist. 23 (1995) 1.
[8] J.M. Cushing, Non-linear matrix models and population dynamics, Nat. Resour. Model. 2 (1988) 539.

[9] J.M. Cushing, Y. Zhou, The net reproductive number and stability in linear structured population models, Nat. Resour. Model. 8 (1994) 297.
[10] V. Favrichon, Modeling the dynamics and species composition of tropical mixed-species uneven-aged natural forest: Effects of alternative cutting regimes, Forest Sci. 44 (1998) 113.
[11] B. Fontez, La modélisation de la dynamique forestiére: étude d'un modèle stochastique, Mémoire de DEA de biostatistique, Université de Montpellier 2, France, 1996.
[12] S. Gourlet-Fleury, G. Cornu, S. Jésel, H. Dessard, J.G. Jourget, L. Blanc, N. Picard, Using models for predicting recovery and assessing tree species vulnerability in logged tropical forests: a case study from French Guiana, Forest Ecol. Manage. 209 (1–2) (2005) 69.
[13] S. Gourlet-Fleury, J.M. Guehl, O. Laroussinie, Ecology and Management of a Neotropical Rainforest, Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana, Elsevier, Paris, 2004, p. 281.
[14] J. van Groenendael, H. de Kroon, H. Caswell, Projection matrices in population biology, Trends Ecol. Evol. 3 (10) (1988) 264.
[15] F.R. Hampel, The influence curve and its role in robust estimation, J. Amer. Statist. Assoc. 69 (1974) 383.
[16] F.R. Hampel, The Approach Based on Influence Functions, Wiley, New York, 1986.
[17] H.T. Hayslett Jr., D.S. Solomon, A matrix model for predicting foliage weight of trees by age classes, Math. Biosci. 67 (1) (1983) 113.
[18] F. Houllier, J.D. Lebreton, D. Pontier, Sampling properties of the asymptotic behavior of age- or stage-grouped population models, Math. Biosci. 95 (1989) 161.
[19] P.J. Huber, Robust Statistics, Wiley, New York, 2004.
[20] L.P. Lefkovitch, The study of population growth in organisms grouped by stages, Biometrics 21 (1965) 1.
[21] P.H. Leslie, In the use of matrices in certain population mathematics, Biometrika 33 (1945) 183.
[22] H.C. Lu, J. Buongiorno, Long- and short-term effects of alternative cutting regimes on economic returns and ecological diversity in mixed-species forests, Forest Ecol. Manage. 58 (3–4) (1993) 173.
[23] B.R. Michie, J. Buongiorno, Estimation of a matrix model of forest growth from re-measured permanent plots, Forest Ecol. Manage. 8 (1984) 127.
[24] J.-M. Naulin, A contribution of sparse matrices tools to matrix population model analysis, Math. Biosci. 177&178 (2002) 25.
[25] J. Pardé, J. Bouchon, Dendomét rie, second ed., ENGREF, Nancy, 1988.
[26] N. Picard, S. Gourlet-Fleury, V. Favrichon, Modelling the forest dynamics at Paracou: the contributions of five models, in: S. Gourlet-Fleury, J.M. Guehl, O. Laroussinie (Eds.), Ecology and Management of a Neotropical Rainforest, Lessons drawn from Paracou, a long-term experimental research site in French Guiana, Elsevier, Paris, 2004, p. 281.
[27] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. Available from: <http://www.R-project.org>.
[28] J. Ripoll, J. Saldaña, J.C. Senar, Evolutionarily stable transition rates in a stage-structured-model. An application to the analysis of size distributions of badges of social status, Math. Biosci. 190 (2004) 145.
[29] T. Takada, H. Nakajima, An analysis of life history evolution in terms of the density-dependent Lefkovitch matrix-model, Math. Biosci. 112 (1) (1992) 155.
[30] M.B. Usher, A matrix approch to the management of renewable resources, with special reference to the selection forests, J. Appl. Ecol. 3 (1966) 355.
[31] M.B. Usher, A matrix model for forest management, Biometrics 25 (1969) 309.
[32] M. Zhou, J. Buongiorno, Nonlinearity and noise interaction in a model of forest growth, Ecol. Model. 180 (2–3) (2004) 291.

Annexe B

Calculs annexes à l'article paru dans *Mathematical Biosciences*

B.1 Calcul des dérivées partielles de λ_1 par rapport aux paramètres

On considère h, l'application qui aux vecteurs des paramètres de U, $\tilde{\theta} = \Pi(\theta)$, associe la première valeur propre λ_1. C'est une application de $\mathbb{R}^{3I-1}$ dans $\mathbb{R}$ définie de façon implicite par : $\Phi(\tilde{\theta}, h(\tilde{\theta})) = 0$, où Φ est une fonction de $\mathbb{R}^{3I-1} \times \mathbb{R}$ dans $\mathbb{R}$, définie par l'équation (2.1). Le théorème des fonctions donne les dérivées partielles de h par rapport aux paramètres $\tilde{\theta}_i$, pour $i = 1, \ldots, 3I-1$:

$$\frac{\partial h}{\partial \tilde{\theta}_i} = -\frac{\frac{\partial \Phi}{\tilde{\theta}_i}}{\frac{\partial \Phi}{\lambda_1}}$$

En développant le déterminant dans le définition de la fonction Φ, on obtient la forme développée de Φ en fonction des paramètres :

$$\Phi(\tilde{\theta}, \lambda_1) = P_0 + \sum_{i=1}^{I} (-1)^{i+1} f_i Q_i P_i$$

avec $q_1 = 1$ et :

$$\begin{cases} Q_i = \prod_{j=1}^{i} q_j & 1 \leq i \leq I \\ P_i = \prod_{k=i+1}^{I} (p_k - \lambda_1) & 0 \leq i \leq I-1 \\ P_I = 1 \end{cases}$$

On calcule la dérivée par rapport aux p_l :

$$\frac{\partial \Phi}{\partial p_l} = \frac{\partial P_0}{\partial p_l} + \sum_{i=1}^{I} (-1)^{i+1} f_i Q_i \frac{\partial P_i}{\partial p_l}$$

avec :

$$\frac{\partial P_i}{\partial p_l} = \begin{cases} \prod_{k=i+1, k\neq l}^{I} (p_k - \lambda_1) = \dfrac{P_i}{p_l - \lambda_1} & \text{si } i \leq l-1 \\ 0 & \text{sinon} \end{cases}$$

puis la dérivée par rapport aux q_l :

$$\frac{\partial \Phi}{\partial q_l} = \sum_{i=1}^{I} (-1)^{i+1} f_i \frac{\partial Q_i}{\partial q_l} P_i$$

avec :

$$\frac{\partial Q_i}{\partial q_l} = \begin{cases} \prod_{j=1, j\neq l}^{i} q_j = \dfrac{Q_i}{q_l} & \text{si } i \geq l \\ 0 & \text{sinon} \end{cases}$$

enfin la dérivée par rapport aux f_l :

$$\frac{\partial \Phi}{\partial f_l} = (-1)^{l+1} Q_l P_l$$

On calcule ensuite la dérivée partielle de Φ par rapport à λ_1

$$\frac{\partial \Phi}{\partial \lambda_1} = \frac{\partial P_0}{\partial \lambda_1} + \sum_{i=1}^{I} (-1)^{i+1} f_i Q_i \frac{\partial P_i}{\partial \lambda_1}$$

avec :

$$\begin{cases} \dfrac{\partial P_i}{\partial \lambda_1} = -\sum_{k=i+1}^{I} \prod_{\{j=i+1, j\neq k\}}^{I} (p_j - \lambda_1) = -P_i \sum_{k=i+1}^{I} \dfrac{1}{p_k - \lambda_1} \quad (i = 0 \ldots I-1) \\ \dfrac{\partial P_I}{\partial \lambda_1} = 0 \end{cases}$$

B.2 Calcul des dérivées partielles de w_1 par rapport aux paramètres

La droite de vecteurs propres associés à λ_1 est engendré par $\omega = (\omega_k)_{k=1,\dots,I}$, dont la première composante est égale à 1, et qui s'écrit, en résolvant le système $U\omega = \lambda_1\omega$:

$$\begin{cases} \omega_1(\tilde{\theta}) = 1 \\ \omega_k(\tilde{\theta}) = \prod_{j=2}^{k} \dfrac{q_j}{\lambda_1(\tilde{\theta}) - p_j} \end{cases}$$

Il est immédiat que, pour tout $i = 1, \dots, 3I-1$:

$$\frac{\partial \omega_1}{\partial \tilde{\theta}_i} = 0$$

D'autre part, le logarithme de ω_k, pour $k \geq 2$ s'écrit :

$$\ln(\omega_k) = \sum_{j=2}^{k} \ln(q_j) - \sum_{j=2}^{k} \ln(\lambda_1 - p_j)$$

De la relation :

$$\frac{\partial \ln(\omega_k)}{\partial .} = \frac{1}{\omega_k}\frac{\partial \omega_k}{\partial .}$$

on en déduit les dérivées du logarithme de ω_k par rapport aux paramètres :

$$\frac{\partial \ln(\omega_k)}{\partial p_l} = \left[-\frac{\partial \lambda_1}{p_l} \sum_{j=2}^{k} \frac{1}{\lambda_1 - p_j} + \chi_l^p \right] \omega_k$$

avec :

$$\chi_l^p = \begin{cases} 0 & \text{si } l > k \text{ et } l = 1 \\ \dfrac{1}{\lambda_1 - p_l} & \text{si } 2 \leq l \leq k \end{cases}$$

puis :

$$\frac{\partial \ln(\omega_k)}{\partial q_l} = \left[-\frac{\partial \lambda_1}{q_l} \sum_{j=2}^{k} \frac{1}{\lambda_1 - p_j} + \chi_l^q \right] \omega_k$$

avec :

$$\chi_l^q = \begin{cases} 0 & \text{si } l > k \\ \dfrac{1}{q_l} & \text{si } 2 \leq l \leq k \end{cases}$$

et enfin :

$$\frac{\partial \ln(\omega_k)}{\partial f_l} = -\frac{\partial \lambda_1}{f_l} \sum_{j=2}^{k} \frac{1}{\lambda_1 - p_j} \omega_k$$

Soit $w_1 = (w_{1\,k})_k$ le vecteur propre colinéaire à ω et de norme 1 :

$$w_1 = \frac{\omega}{\|\omega\|}$$

Ainsi :

$$\begin{aligned} \frac{\partial w_{1\,k}}{\partial .} &= \frac{1}{\|\omega\|} \frac{\partial \omega_k}{\partial .} - \frac{1}{\|\omega\|^2} \frac{\partial \|\omega\|}{\partial .} \omega_k \\ &= \left(\frac{\partial \omega_k}{\partial .} - \frac{\partial \|\omega\|}{\partial .} w_{1\,k} \right) \frac{1}{\|\omega\|} \end{aligned}$$

Or :

$$\begin{aligned} \frac{\partial \|\omega\|}{\partial .} &= \frac{1}{2\|\omega\|} \frac{\partial \|\omega\|^2}{\partial .} \\ &= \frac{1}{\|\omega\|} \sum_{j=1}^{n} \frac{\partial \omega_j}{\partial .} \omega_j \end{aligned}$$

Donc :

$$\frac{\partial w_{1\,k}}{\partial .} = \left(\frac{1}{\omega_k} \frac{\partial \omega_k}{\partial .} - \frac{1}{\|\omega\|^2} \sum_{j=1}^{n} \frac{\partial \omega_j}{\partial .} \omega_j \right) w_{1\,k}$$

On en déduit ainsi les dérivées partielles de w_1 par rapport aux paramètres. Elles dépendant de λ_1 et des dérivées partielles de λ_1 par rapport aux paramètres calculés plus haut.

Annexe C

Article sous presse dans *Computational Statistics and Data Analysis*

Robustness of the estimators of transition rates for size-classified matrix models

M. Zetlaoui [a] N. Picard [b] A. Bar-Hen [c,*]

[a] *Institut National Agronomique Paris-Grignon, 16 rue Claude Bernard, 75231 Paris Cedex 5, France*

[b] *Cirad, forest department, Campus de Baillarguet, TA 10/D, 34398 Montpellier Cedex 5, France*

[c] *INA-PG OMIP, 16, rue Claude Bernard, F-75231 Paris cedex 5 & Université Paris 13, SMBH-Lim & Bio, 74, rue Marcel Cachin, 93017 Bobigny Cedex, France*

Abstract

Matrix models are often used to model the dynamics of age-structured or size-structured populations. The Usher model is an important particular case that rely on the following hypothesis: between time steps t and $t+1$, individuals either remain in the same class, move up to the following class, or die. There are then two ways of handling data that do not meet this condition: either remove them prior to data analysis, or rectify them. These two ways correspond to two estimators of transition parameters. The former, that corresponds to the classical estimator, is obtained from the latter by a data trimming. The two estimators of transition parameters are compared on the basis of their robustness in order to obtain a criterion of choice between the two estimators. The influence curve of both estimators is computed, then their gross sensitivity and their asymptotic variance. The untrimmed estimator is more robust than the classical one. Its asymptotic variance can be lower or greater than that of the classical estimator depending on the boundaries used for data trimming. The results are applied to a tropical rain forest in French Guiana, with a discussion on the role of the class width.

Key words: Influence curve; Matrix model; Parameter estimation; Robustness; Usher model

* Corresponding author: A. Bar-Hen, INA-PG OMIP, 16, rue Claude Bernard, F-75231 Paris cedex 5

Email addresses: melanie.zetlaoui@libertysurf.fr (M. Zetlaoui), nicolas.picard@cirad.fr (N. Picard), avner.bar-hen@smbh.univ-paris13.fr (A. Bar-Hen).

1 Introduction

Models of population dynamics are an important tool in many ecological studies. They are used to mimic the future evolution of the population. Among the discrete-time models, matrix model are often used to study the dynamics of structured populations (either age-structured or size-structured populations). They also permit to simplify the dynamics of a population into its basic components: recruitment or birth, growth or ageing, and mortality. The Usher model is a matrix model for size-structured populations that is based on four hypothesis (Usher, 1966, 1969):

- Hypothesis of independence: the evolution of the individuals are independent
- Markov hypothesis: the evolution of an individual between two time steps t and $t+1$ depends only on its state at time t
- Usher hypothesis: during each time step, an individual can stay in the same class, move up a class, or die; each individual may give birth to a number of offspring
- Hypothesis of stationarity: the evolution of individuals between two time step is independent of time.

A more general matrix model was proposed by Lefkovitch (1965), and allows any transition from one stage to another. The Leslie (1945) model, which describes a population grouped by age, is a special case of the Usher model: during each time step, an individual can only move up a class or die.

In this paper, we are interested with the Usher model and particularly with the Usher hypothesis. This hypothesis reduces the number of parameters of the model, compared with a more general matrix model such as the Lefkovitch model. When facing limited data, it is thus a useful alternative that permits more precise parameter estimates. The only restriction is that the dynamic has to match the constraints imposed by the hypothesis. This is generally the case for size-structured populations where growth can be described by a first-order process. On the contrary, this is rarely the case for stage-structured populations. The trick then is to choose classes and time step so that Usher hypothesis is realistic (Favrichon, 1998). These reasons can explain in particular why Usher models have been intensely used in forestry to model the dynamics of diameter distributions.

Nevertheless, even when using proper classes and time step, some data can violate Usher hypothesis. We consider here two ways of handling these data:

- Remove them prior to data analysis
- Rectify them: individuals that move up by more than one class in a time step are considered as individuals that move up a single class, and individuals

that move backwards are considered as individuals that stay in the same class.

We thus obtain two versions of transition parameters depending on whether data are trimmed or not.

The purpose of this study is to present a criterion of choice between these two data treatments, based on the robustness of the estimator of transition parameters. To compare the robustness of these two estimators of transition parameters, we use the influence curve. This gives a quantitative criterion of comparison of robustness of estimators . Furthermore, it gives the asymptotic behavior of an estimator. We compare then the asymptotic variance of the estimators and the difference between their expectations. Finally, the results are applied to a data set obtained from experimental forest plots in a tropical rain forest at Paracou (French Guiana).

2 Influence function

The approach to robust estimation of a location parameter by the influence curve was developed by Hampel (1974). It deals with estimators viewed as functionals. Let $(X_1, \ldots, X_n)$ be a n-sample with distribution function $F(\theta)$ on $\mathbf{R}^d$, with θ an unknown parameter. θ is supposed to express itself in the form $T(F)$. Then, the empirical estimator of θ is $T(F_n)$, where F_n is the empirical distribution function. The estimator is said to be generated by the functional T. Then, the influence curve, $IC_{T,F}(x)$, defined, under conditions of differentiability of T, by :

$$IC_{T,F}(x) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon},$$

where δ_x is a Dirac mass at x, describes the effect of an additional observation $x \in \mathbf{R}^d$ on the statistic $T(F)$. It gives the asymptotic behavior of the estimator:

$$\sqrt{n}(T(F_n) - T(F)) \to \mathcal{N}\left(0, \int IC_{T,F}(x)^2 dF(x)\right),$$

where the convergence is in law. Furthermore, the influence curve permits to obtain criteria to assess the robustness of an estimator. A first criterion is given by:

$$\mu = \sup_x \|IC_{T,F}(x)\|_\infty,$$

where $\|.\|_\infty$ is the infinity norm. μ has been called the gross error sensitivity by Hampel (1974). It measures the biggest influence caused by a little perturbation on the value of the estimator.

A second criterion deals with the little fluctuations on the observations. The local sensitivity is defined by:

$$\nu = \sup_{x \neq y} \frac{\|IC_{T,F}(y) - IC_{T,F}(x)\|_\infty}{\|y - x\|_\infty}.$$

It measures the effect on the statistic of the replacement of an observation x with another observation y. Particularly, it measures the effect of regrouping on the statistic.

3 The Usher model

3.1 Mathematical model

The Usher model lies on the description of the evolution of the population by a vector, $N(t)$, of the number of individuals in I ordered state classes at time t. Its components, $N_i(t)$ for $i = 1, \ldots, I$, are the number of individuals in each class i. Time is discrete. The relation between $N(t)$ and $N(t+1)$ is given by a $I \times I$ transition matrix A, called the Usher matrix (Usher, 1966, 1969):

$$N(t+1) = A\,N(t).$$

The matrix A can be written as $A = PS + R$, where:

$$P = \begin{pmatrix} 1-q_1 & & 0 & 0 \\ q_1 & \ddots & & \vdots \\ & \ddots & 1-q_{I-1} & 0 \\ 0 & & q_{I-1} & 1 \end{pmatrix}, \; S = \begin{pmatrix} 1-m_1 & & 0 \\ & \ddots & \\ & & \ddots & \\ 0 & & & 1-m_I \end{pmatrix}$$

$$R = \begin{pmatrix} f & \ldots & f \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{pmatrix}.$$

q_i is the conditional transition probability for an individual in state i at time t to move up to state $i+1$ at time $t+1$ knowing that it stays alive, m_i is the probability that an individual dies in state i, and f in the average fecundity. q_i and m_i take values in $]0,1[$, whereas f takes values in $\mathbf{R}^+$. Thus P describe growth, S is the survival matrix, and R describe regeneration.

Only the diagonal and the sub-diagonal of P contains non null elements due to Usher hypothesis. Death probabilities m_i are estimated from Boolean events (dead / alive) and the fecundity is estimated from count data, which limits the form of their estimators. On the contrary, transition probabilities q_i are potentially estimated from continuous size variables, and its estimators can take various forms. Hence, we focus on the estimation of the parameters of the matrix P.

We are interested with the robustness of the estimators of q_i because of measurements errors, that result in misclassifications of the individuals in stages, and thus influence the estimates of the q_i. To model misclassifications of individuals in stages, it is necessary to turn back to the continuous size variables that are used for classification. So we also have to specify a statistical model for size variables.

3.2 Statistical model

Hereafter, a class is an interval of values for a continuous size variable X. We consider a class $C = [l_0; l_1[$. Let X the size at time t of an individual in class C that stays alive between time t and $t+1$, and ΔX its size increment between these two time steps . We suppose that:

- X follows a law with density f and distribution function F. The function f is defined on $[l_0; l_1[$.
- the law of $\Delta X|X$ is defined by the density $g(.|X)$ and the distribution function G_X. For any size x, the support of the function $y \mapsto g(y|x)$ is equal to $[-x; +\infty[$.

The law of the pair $(X, \Delta X)$ knowing that $X \in C$, denoted H, has the density:

$$h(x, y) = g(y|x)f(x).$$

Two versions of the transition probability q for class C are considered in this study, one of which is the classical expression for q. The classical transition probability, denoted $q_{MM'}$, relies on a trim of the continuous size variable X prior to parameter estimation. This trim is applied to ensure that Usher hypothesis is verified. It is defined as the conditional probability for an individual in class C at time t to move up to the next class at time $t+1$ knowing that its size at time $t+1$ is less than a positive real M and greater than a positive real M'. The thresholds M and M' are classically chosen to ensure that the data meet Usher hypothesis. We suppose that: $0 \leq M' \leq l_0$ and $l_1 < M$.

The other version of q that we consider does not trim data. The data used for parameter estimation do not have to meet Usher hypothesis, and the transition probability is defined as the probability for an individual in class C at time t to overcome l_1 in size at time $t+1$. It should be clear that overcoming l_1 in size is not equivalent with moving up to the class following C, as this class, whose lower bound is l_1, has an upper bound. This second probability is denoted q_∞ and is the limit of $q_{MM'}$ when M' is equal to zero and M tends to infinity. Those probabilities are written as:

$$q_{MM'} = \Pr(l_1 \leq X + \Delta X | X \in C \text{ and } M' \leq X + \Delta X < M), \quad (1)$$
$$q_\infty = \Pr(l_1 \leq X + \Delta X | X \in C). \quad (2)$$

We compare the robustness of the empirical estimators of both transition probabilities using the criteria given by the influence curve. As a corollary, the asymptotic variance of the estimators and the difference between their expectations are also given.

4 Robustness of estimators of transition probability

In this part, we study the robustness of the two proposed estimators of transition probability. At first, we express the functionals which generated them. Then, we calculate their influence curve in order to calculate their sensibility and their asymptotic variance.

4.1 Estimators of transition probability

Given a n-sample $(X_i, \Delta X_i)$ of independent pairs that are identically distributed according to H, the empirical estimators of $q_{MM'}$ and q_∞ are:

$$\hat{q}_{MM'} = \frac{\sum_{i=1}^n \mathbf{1}(l_1 < X_i + \Delta X_i < M)}{\sum_{i=1}^n \mathbf{1}(M' \leq X_i + \Delta X_i < M)},$$
$$\hat{q}_\infty = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(l_1 < X_i + \Delta X_i).$$

Let $T_{MM'}$ and T be the functionals which generate the empirical estimators of $q_{MM'}$ and q_∞ respectively, defined by the equations:

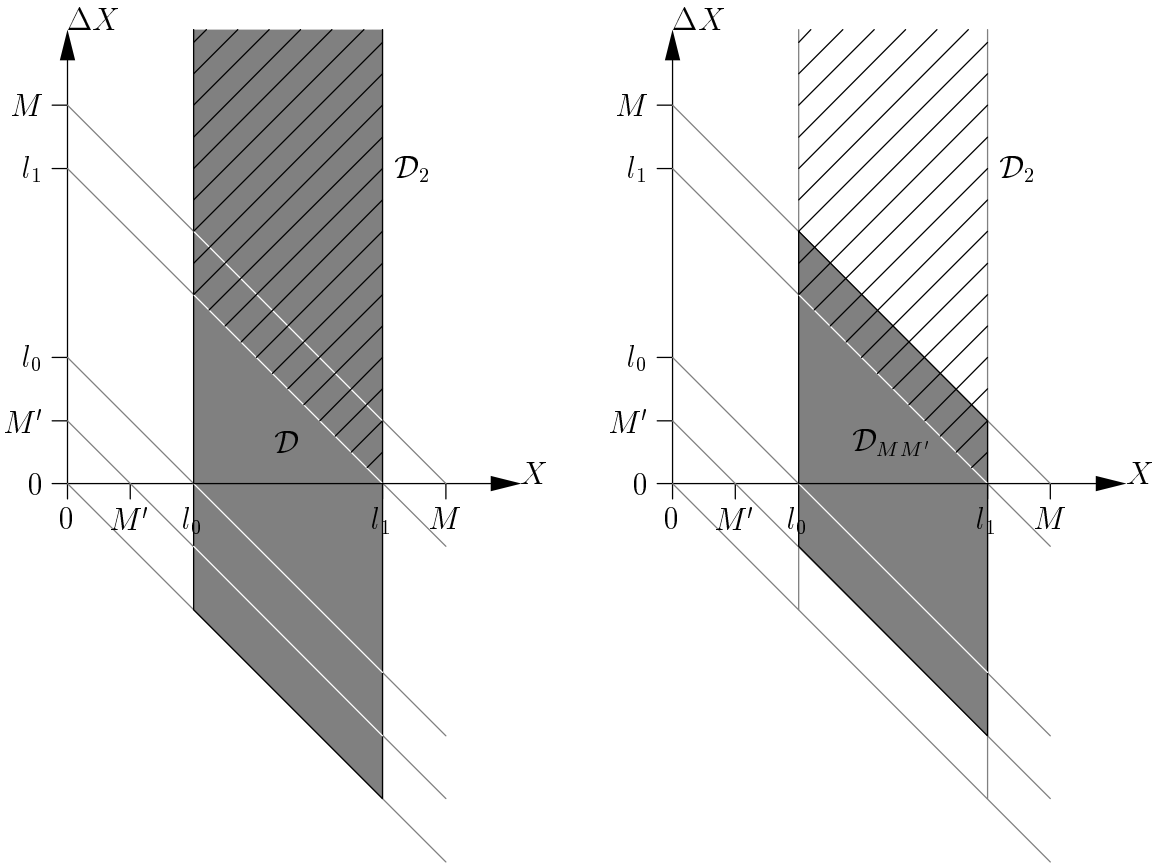


Fig. 1. Domains used to define the probabilities q_∞ (left figure) and $q_{MM'}$ (right figure). The dashed domain is $\mathcal{D}_2$. The grey domain is $\mathcal{D}$ (left figure) or $\mathcal{D}_{MM'}$ (right figure).

$$T_{MM'}(H) = \frac{\iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} h(x,y)\mathrm{d}x\mathrm{d}y}{\iint_{\mathcal{D}_{MM'}} h(x,y)\mathrm{d}x\mathrm{d}y}, \tag{3}$$

$$T(H) = \iint_{\mathcal{D}_2} h(x,y)\mathrm{d}y\mathrm{d}x, \tag{4}$$

where the domains are (figure 1):

$$\mathcal{D} = \{(x,y); l_0 \leq x < l_1, x+y \geq 0\},$$
$$\mathcal{D}_2 = \{(x,y) \in \mathcal{D}; x+y \geq l_1\},$$

and:

$$\mathcal{D}_M = \{(x,y) \in \mathcal{D}; x+y < M\},$$
$$\mathcal{D}_{M'} = \{(x,y) \in \mathcal{D}; x+y, \geq M'\},$$
$$\mathcal{D}_{MM'} = \mathcal{D}_M \cap \mathcal{D}_{M'}.$$

We denote also $\mathcal{D}_1$ the complementary of $\mathcal{D}_2$ in $\mathcal{D}$, defined by:

$$\mathcal{D}_1 = \{(x, y) \in \mathcal{D}; x + y < l_1\}.$$

The functional $T_{MM'}$ can be relied to the functional T. Indeed, we put:

$$\begin{aligned}\alpha_{MM'} &= \iint_{\mathcal{D}\setminus\mathcal{D}_{MM'}} h(x, y)\mathrm{d}x\mathrm{d}y \\ &= 1 - \iint_{\mathcal{D}_{MM'}} h(x, y)\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{5}$$

and:

$$\begin{aligned}\alpha_M &= 1 - \iint_{\mathcal{D}_M} h(x, y)\mathrm{d}x\mathrm{d}y, \\ \alpha_{M'} &= 1 - \iint_{\mathcal{D}_{M'}} h(x, y)\mathrm{d}x\mathrm{d}y.\end{aligned}$$

When M equals the upper bound of the class that follows C and $M' = l_0$, $\alpha_{MM'}$ is the proportion of data that violate Usher's hypothesis. Particularly, $\alpha_{MM'}$ verifies: $0 \leq \alpha_{MM'} \leq 1$, and: $\alpha_{MM'} = \alpha_M + \alpha_{M'}$. The case $\alpha_{MM'} = 1$ corresponds to the case where all individuals have their size at time $t + 1$ greater than M or between zero and M'. Afterwards, we do not consider this case, and we suppose that: $0 \leq \alpha_{MM'} < 1$. Furthermore, as:

$$T(H) = \iint_{\mathcal{D}_{MM'}\cap\mathcal{D}_2} h(x, y)\mathrm{d}x\mathrm{d}y + \iint_{\mathcal{D}_2\setminus(\mathcal{D}_{MM'}\cap\mathcal{D}_2)} h(x, y)\mathrm{d}x\mathrm{d}y.$$

We obtain:

$$T(H) = (1 - \alpha_{MM'})T_{MM'}(H) + \alpha_M, \tag{6}$$

which is equivalent to:

$$T_{MM'}(H) = \frac{T(H) - \alpha_M}{1 - \alpha_{MM'}}. \tag{7}$$

Furthermore, this last equation implies that:

$$T(H) \geq \alpha_M.$$

4.2 Influence of the parameters

In this part, we calculate the influence curve of the functionals $T_{MM'}$ and T. Let (u, v) in $\mathcal{D}$. The expression of the influence curve of $T_{MM'}$ at point H in direction (u, v), $IC_{T_{MM'},H}$, is given by the following theorem

Proposition 1 *The influence curve of the trimmed estimator generated by the functional $T_{MM'}$, at point of the distribution of individual size vector, H, in direction of the point (u,v) in $\mathcal{D}$, $IC_{T_{MM'},H}$, is equal to:*

$$IC_{T_{MM'},H}((u,v)) = \begin{cases} -\dfrac{T(H)-\alpha_M}{(1-\alpha_{MM'})^2} & \text{for } (u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_1 \\ \dfrac{1-\alpha_{M'}-T(H)}{(1-\alpha_{MM'})^2} & \text{for } (u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2 \\ 0 & \text{for } (u,v) \in \mathcal{D} \setminus \mathcal{D}_{MM'}, \end{cases}$$

where the domains $\mathcal{D}$, $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_{MM'}$, and the proportion $\alpha_{MM'}$, α_M and $\alpha_{M'}$ are defined in the previous paragraph.

Proof: The influence curve of $T_{MM'}$ at point H in direction $(u,v) \in \mathcal{D}$, $IC_{T_{MM'},H}$ is written as:

$$\begin{aligned}
&IC_{T_{MM'},H}((u,v)) \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\frac{\iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} (1-\epsilon) h(x,y) \mathrm{d}x\mathrm{d}y + \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2}}{\iint_{\mathcal{D}_{MM'}} (1-\epsilon) h(x,y) \mathrm{d}x\mathrm{d}y + \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'}}} - \frac{\iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} h(x,y) \mathrm{d}x\mathrm{d}y}{\iint_{\mathcal{D}_{MM'}} h(x,y) \mathrm{d}x\mathrm{d}y} \right] \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \frac{\epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2} \iint_{\mathcal{D}_{MM'}} h(x,y) \mathrm{d}x\mathrm{d}y - \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'}} \iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} h(x,y) \mathrm{d}x\mathrm{d}y}{\left(\iint_{\mathcal{D}_{MM'}} (1-\epsilon) h(x,y) \mathrm{d}x\mathrm{d}y + \epsilon \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'}} \right) \iint_{\mathcal{D}_{MM'}} h(x,y) \mathrm{d}x\mathrm{d}y} \\
&= \frac{\mathbf{1}_{(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2} \iint_{\mathcal{D}_{MM'}} h(x,y) \mathrm{d}x\mathrm{d}y - \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'}} \iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} h(x,y) \mathrm{d}x\mathrm{d}y}{\left(\iint_{\mathcal{D}_{MM'}} h(x,y) \mathrm{d}x\mathrm{d}y \right)^2} \\
&= \frac{\mathbf{1}_{(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2} - \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'}} T_{MM'}(H)}{\iint_{\mathcal{D}_{MM'}} h(x,y) \mathrm{d}x\mathrm{d}y} \\
&= \frac{\mathbf{1}_{(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2} - \mathbf{1}_{(u,v) \in \mathcal{D}_{MM'}} T_{MM'}(H)}{1-\alpha_{MM'}},
\end{aligned}$$

where the last equality uses equation (5). When $(u,v) \notin \mathcal{D}_{MM'}$, this simplifies to zero. As $\mathcal{D}_1$ and $\mathcal{D}_2$ do not overlap, nor do $\mathcal{D}_{MM'} \cap \mathcal{D}_1$ and $\mathcal{D}_{MM'} \cap \mathcal{D}_2$. When $(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_1$, the expression of $IC_{T_{MM'},H}((u,v))$ thus simplifies to: $-T_{MM'}(H)/(1-\alpha_{MM'})$, whereas when $(u,v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2$, it simplifies to: $(1-T_{MM'}(H))/(1-\alpha_{MM'})$. Replacing $T_{MM'}(H)$ by its expression (7) that depends on $T(H)$ finally gives the Proposition. ∎

We remark that the influence of the functional $T_{MM'}$ is constant on two contiguous ensembles: $\mathcal{D}_1 \cap \mathcal{D}_{MM'}$ and $\mathcal{D}_2 \cap \mathcal{D}_{MM'}$. Consequently, the local sensitivities of $T_{MM'}$ is infinite.

We deduce from the previous proposition the influence curve of T at point H in direction (u,v), $IC_{T,H}$, given by the:

Corollary 1 *The influence curve of the untrimmed estimator generated by*

the functional T, at point of the distribution of individual size vector, H, in direction of the point (u, v) in $\mathcal{D}$, $IC_{T,H}$, is equal to:

$$IC_{T,H}((u,v)) = \begin{cases} -T(H) & \text{for } (u,v) \in \mathcal{D}_1 \\ 1 - T(H) & \text{for } (u,v) \in \mathcal{D}_2, \end{cases}$$

where the domains $\mathcal{D}$, $\mathcal{D}_1$ and $\mathcal{D}_2$ are defined in the previous paragraph.

Proof: Let (u, v) a pair in $\mathcal{D}$. We set M' to zero. The domain $\mathcal{D}_{MM'}$ is then confounded with $\mathcal{D}_M$, which implies $\alpha_{MM'} = \alpha_M$. To simplify notations, we also denote T_M in place of $T_{MM'}$ when $M' = 0$.

At first, we show than $IC_{T,H}((u, v))$ is the simple limit of $IC_{T_M,H}((u, v))$ when M tends to infinity, and we deduce after the expression of $IC_{T,H}((u, v))$.

By definition, $IC_{T,H}((u, v))$ is equal to:

$$IC_{T,H}((u,v)) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)H + \epsilon\delta(u,v)) - T(H)}{\epsilon}.$$

Now, T_M converge uniformly towards T as $M \to \infty$. Indeed, by the equation (7):

$$T_M(H) - T(H) = \frac{\alpha_M}{1 - \alpha_M}(T(H) - 1). \tag{8}$$

Then:

$$\|T_M - T\|_\infty \leq \frac{\alpha_M}{1 - \alpha_M},$$

where $\|.\|_\infty$ is the infinity norm in the space of bivariate distributions. Now:

$$\alpha_M = \iint_{(x,y)\in\mathcal{D}} h(x,y)\mathrm{d}x\mathrm{d}y - \iint_{(x,y)\in\mathcal{D},\, x+y\leq M} h(x,y)\mathrm{d}x\mathrm{d}y,$$

and:

$$\lim_{M\to\infty} \iint_{(x,y)\in\mathcal{D},\, x+y\leq M} h(x,y)\mathrm{d}x\mathrm{d}y = \iint_{(x,y)\in\mathcal{D}} h(x,y)\mathrm{d}x\mathrm{d}y,$$

because h is integrable in $\mathcal{D}$, therefore:

$$\lim_{M\to\infty} \alpha_M = 0. \tag{9}$$

Then:

$$\lim_{M\to\infty} \|T_M - T\|_\infty = 0,$$

and T_M converge uniformly towards T. We deduce that:

$$
\begin{aligned}
IC_{T,H}((u,v)) &= \lim_{\epsilon\to 0}\lim_{M\to\infty}\frac{T_M((1-\epsilon)H+\epsilon\delta(u,v))-T_M(H)}{\epsilon}\\
&= \lim_{M\to\infty}\lim_{\epsilon\to 0}\frac{T_M((1-\epsilon)H+\epsilon\delta(u,v))-T_M(H)}{\epsilon}\\
&= \lim_{M\to\infty} IC_{T_M,H}((u,v)).
\end{aligned}
$$

Proposition 1 gives the expression of $IC_{T_M,H}((u,v))$ by setting $M'=0$:

$$
IC_{T_M,H}((u,v)) = -\frac{T(H)-\alpha_M}{(1-\alpha_M)^2}\mathbf{1}_{(u,v)\in\mathcal{D}_1} + \frac{1-T(H)}{(1-\alpha_M)^2}\mathbf{1}_{(u,v)\in\mathcal{D}_M\cap\mathcal{D}_2}.
$$

Now:

$$
\lim_{M\to\infty}\mathbf{1}_{(u,v)\in\mathcal{D}_M} = 1. \tag{10}
$$

Indeed, as (u,v) is in $\mathcal{D}$, then:

$$
\exists\widetilde{M}>0;\forall M\geq\widetilde{M}\quad u+v\leq M,
$$

and:

$$
\exists\widetilde{M}>0;\forall M\geq\widetilde{M}\quad \mathbf{1}_{(u,v)\in\mathcal{D}_M}=1.
$$

Then, we deduce from (9) and (10) the expression of $IC_{T,H}$ given in the corollary 1. ■

As for the influence curve of T_M, the influence of the functional T is constant on two contiguous ensembles: $\mathcal{D}_1$ and $\mathcal{D}_2$. Consequently, the local sensitivities of T is also infinite. In the following paragraph, we compare the gross sensitivities of T and T_M.

4.3 Sensitivity of the parameters

Proposition 1 and Corollary 1 give immediately the gross sensitivities $\gamma_{MM'}$ and γ of $T_{MM'}$ and T respectively:

$$
\begin{aligned}
\gamma_{MM'} &= \max\left\{\frac{T(H)-\alpha_M}{(1-\alpha_{M'}-\alpha_M)^2};\frac{1-\alpha_{M'}-T(H)}{(1-\alpha_{M'}-\alpha_M)^2}\right\},\\
\gamma &= \max\{T(H);1-T(H)\}.
\end{aligned}
$$

We have the following result:

Proposition 2 $\gamma_{MM'}\geq\gamma$.

Proof: We have:

$$
\gamma_{MM'}\geq\gamma\Leftrightarrow\max\left\{\frac{T(H)-\alpha_M}{(1-\alpha_{M'}-\alpha_M)^2};\frac{1-\alpha_{M'}-T(H)}{(1-\alpha_{M'}-\alpha_M)^2}\right\}\geq\max\{T(H);1-T(H)\},
$$

that is:

$$\gamma_{MM'} \geq \gamma \Leftrightarrow \begin{cases} \left(\dfrac{T(H) - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2} \geq T(H) \text{ and } \dfrac{T(H) - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2} \geq 1 - T(H) \right) \\ \text{or} \\ \left(\dfrac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{M'} - \alpha_M)^2} \geq T(H) \text{ and } \dfrac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{M'} - \alpha_M)^2} \geq 1 - T(H) \right). \end{cases}$$

First of all, we compare $\dfrac{T(H) - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2}$ with $T(H)$ and $1 - T(H)$. First, we have:

$$\frac{T(H) - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2} - T(H) = \frac{T(H)\left[1 - (1 - \alpha_{M'} - \alpha_M)^2\right] - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2}.$$

Then:

$$\frac{T(H) - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2} \geq T(H) \Leftrightarrow T(H) \geq \frac{\alpha_M}{1 - (1 - \alpha_{M'} - \alpha_M)^2}.$$

In the same way, we have:

$$\frac{T(H) - \alpha_M}{(1 - \alpha_{M'} - \alpha_M)^2} \geq 1 - T(H) \Leftrightarrow T(H) \geq \frac{\alpha_M + (1 - \alpha_{M'} - \alpha_M)^2}{1 + (1 - \alpha_{M'} - \alpha_M)^2}.$$

Furthermore, we compare likewise $\dfrac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{M'} - \alpha_M)^2}$ with $T(H)$ and $1 - T(H)$. We obtain:

$$\frac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{M'} - \alpha_M)^2} \geq T(H) \Leftrightarrow T(H) \leq \frac{1 - \alpha_{M'}}{1 + (1 - \alpha_{M'} - \alpha_M)^2},$$

and:

$$\frac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{M'} - \alpha_M)^2} \geq 1 - T(H) \Leftrightarrow T(H) \leq \frac{1 - \alpha_{M'} - (1 - \alpha_{M'} - \alpha_M)^2}{1 - (1 - \alpha_{M'} - \alpha_M)^2}.$$

Finally:

$$\gamma_{MM'} \geq \gamma \Leftrightarrow \begin{cases} \max\left\{ \dfrac{\alpha_M}{1 - (1 - \alpha_{M'} - \alpha_M)^2}; \dfrac{\alpha_M + (1 - \alpha_{M'} - \alpha_M)^2}{1 + (1 - \alpha_{M'} - \alpha_M)^2} \right\} \leq T(H) \\ \text{or} \\ T(H) \leq \min\left\{ \dfrac{(1 - \alpha_{M'}) - (1 - \alpha_{M'} - \alpha_M)^2}{1 - (1 - \alpha_{M'} - \alpha_M)^2}; \dfrac{1 - \alpha_{M'}}{1 + (1 - \alpha_{M'} - \alpha_M)^2} \right\}. \end{cases}$$

Now, we have directly (since $0 \leq \alpha_M + \alpha_{M'} \leq 1$):

$$\frac{\alpha_M}{1-(1-\alpha_{M'}-\alpha_M)^2} \leq \frac{1-\alpha_{M'}-(1-\alpha_{M'}-\alpha_M)^2}{1-(1-\alpha_{M'}-\alpha_M)^2}.$$

With a inequality resolution, we show that:

$$\frac{\alpha_M}{1-(1-\alpha_{M'}-\alpha_M)^2} \leq \frac{1-\alpha_{M'}}{1+(1-\alpha_{M'}-\alpha_M)^2}.$$

Then, we have:

$$\frac{\alpha_M}{1-(1-\alpha_{M'}-\alpha_M)^2} \leq \min\left\{\frac{1-\alpha_{M'}-(1-\alpha_{M'}-\alpha_M)^2}{1-(1-\alpha_{M'}-\alpha_M)^2}; \frac{1-\alpha_{M'}}{1+(1-\alpha_{M'}-\alpha_M)^2}\right\}.$$

In the same way, we show that:

$$\frac{\alpha_M+(1-\alpha_{M'}-\alpha_M)^2}{1+(1-\alpha_{M'}-\alpha_M)^2} \leq \min\left\{\frac{1-\alpha_{M'}-(1-\alpha_{M'}-\alpha_M)^2}{1-(1-\alpha_{M'}-\alpha_M)^2}; \frac{1-\alpha_{M'}}{1+(1-\alpha_{M'}-\alpha_M)^2}\right\}.$$

Then:

- if $T(H) \leq \min\left\{\frac{1-\alpha_{M'}-(1-\alpha_{M'}-\alpha_M)^2}{1-(1-\alpha_{M'}-\alpha_M)^2}; \frac{1-\alpha_{M'}}{1+(1-\alpha_{M'}-\alpha_M)^2}\right\}$, then $\gamma_{MM'} \geq \gamma$.
- else $T(H) \geq \max\left\{\frac{\alpha_M}{1-(1-\alpha_{M'}-\alpha_M)^2}; \frac{\alpha_M+(1-\alpha_{M'}-\alpha_M)^2}{1+(1-\alpha_{M'}-\alpha_M)^2}\right\}$ and $\gamma_{MM'} \geq \gamma$.

As a conclusion: $\gamma_{MM'} \geq \gamma$. ■

This result indicates that the empirical estimator of the classical transition probability, $T_{MM'}$, is less robust than the untrimmed estimator, T.

We have compared the robustness of empirical estimators of two transition probabilities. The definition of the transition probabilities follows from that of the mathematical model. Then it is also linked to the choice of the state classes.

4.4 Asymptotic variance

In this section we compare the asymptotic variance of both estimators and their expectations. The asymptotic variance of T and $T_{MM'}$, obtained from their influence curve, is equal to respectively:

$$V = \iint_{\mathcal{D}_1} T(H)^2 h(x,y)\mathrm{d}y\mathrm{d}x + \iint_{\mathcal{D}_2} (1-T(H))^2 h(x,y)\mathrm{d}y\mathrm{d}x,$$
$$V_{MM'} = \iint_{\mathcal{D}_{M'}\cap\mathcal{D}_1} \frac{(T(H)-\alpha_M)^2}{(1-\alpha_{MM'})^4} h(x,y)\mathrm{d}y\mathrm{d}x + \iint_{\mathcal{D}_M\cap\mathcal{D}_2} \frac{(1-\alpha_{M'}-T(H))^2}{(1-\alpha_{MM'})^4} h(x,y)\mathrm{d}y\mathrm{d}x,$$

which is equivalent to:

$$V = T(H)(1-T(H)),$$
$$V_{MM'} = \frac{(T(H)-\alpha_M)\,(1-\alpha_{M'}-T(H))}{(1-\alpha_{MM'})^3}.$$

The difference of the two asymptotic variance shows that its sign depends on $T(H)$ and $\alpha_{MM'}$, but this condition is not simple.

On the other hand, the untrimmed estimator $\hat{q}_\infty$ is unbiased, while the trimmed estimator $\hat{q}_{MM'}$ is biased. Nevertheless, they are both asymptotically unbiased. Furthermore, Equation (7) gives the difference between the asymptotic expectation of both estimators:

$$E(\hat{q}_{MM'}) - E(\hat{q}_\infty) = \frac{\alpha_{M'}T(H) + \alpha_M(T(H)-1)}{1-\alpha_{MM'}}.$$

The sign of the difference depends on whether $T(H)$ is greater or less than $\alpha_M/\alpha_{MM'}$. In the particular case where $M'=0$, $E(\hat{q}_{MM'}) \leq E(\hat{q}_\infty)$.

The untrimmed estimator is more robust than the trimmed estimator, but its asymptotic variance can be bigger.

5 Application to the Paracou forest

5.1 Data set and model

Data come from the Paracou experimental site in French Guiana (5°15'N, 52°55'W) (Gourlet-Fleury et al., 2004). It consists of twelve 6.25 ha plots of natural rain forest. The plots were created in 1984. All trees with a diameter at breast height greater than 10 cm, including ingrowth, are monitored. The spatial coordinates and the species of a tree is recorded when it is recruited. Its diameter is measured annually (every two years since 1995), unless it dies. In 1987, nine plots underwent silvicultural treatments with a varying logging intensity. The three remaining plots were left as control.

For this study, we used the diameter data in 1984 and 1986 for all trees that remained alive from 1984 to 1986. The data are thus representative of an

Table 1
Estimates of the parameters of the joint distribution of diameters and diameter increments at Paracou. Standard deviations were computed from the diagonal of an estimate of the Fisher information matrix.

Parameter	Unit	Value	Std. dev.
ξ	cm^{-1}	0.0855	$4\ 10^{-4}$
a_0	cm	−0.2055	$7.4\ 10^{-3}$
a_1	—	0.0286	$6.5\ 10^{-4}$
a_2	cm^{-1}	−0.0004	$1.1\ 10^{-5}$
b_0	cm	0.1322	$2.9\ 10^{-3}$
b_1	—	0.0124	$1.7\ 10^{-4}$

undisturbed forest. The time step is then equal to two years. This gave a sample of $N = 45\,732$ individual vectors $(X, \Delta X)$.

Explanatory analysis showed that the diameter distribution at Paracou in 1984 was well approximated by an exponential distribution:

$$f(x) = \xi \exp[-\xi(x - x_0)],$$

where $x_0 = 10$ cm is the minimum diameter for inventory. The conditional 1984–86 increment distribution knowing the diameter was found to be lognormal. More precisely, diameter increments were minored by -1 cm and the distribution of $\Delta X + 1$ was approximated by a lognormal distribution with parameter $\mu(x)$ and $\sigma(x)$. The relationship between the parameter $\mu(x)$ and diameter x was found to be a second order polynomial whereas the relationship between the parameter $\sigma(x)$ and diameter x was found to be linear:

$$\begin{aligned} g(y|x) &= \frac{1}{\sqrt{2\pi}\sigma(x)(y+1)} \exp\left(-\frac{(\ln(y+1-\mu(x)))^2}{2\sigma(x)^2}\right), \\ \mu(x) &= a_0 + a_1 x + a_2 x^2, \\ \sigma(x) &= b_0 + b_1 x. \end{aligned}$$

The parameter ξ and the vector of parameters of the lognormal distribution, $\theta = (a_0, a_1, a_2, b_0, b_1)$, were estimated by maximizing the log-likelihood of the sample of $(X, \Delta X)$. The maximum likelihood estimates of parameters ξ and θ are given in Table 1.

5.2 Analyses

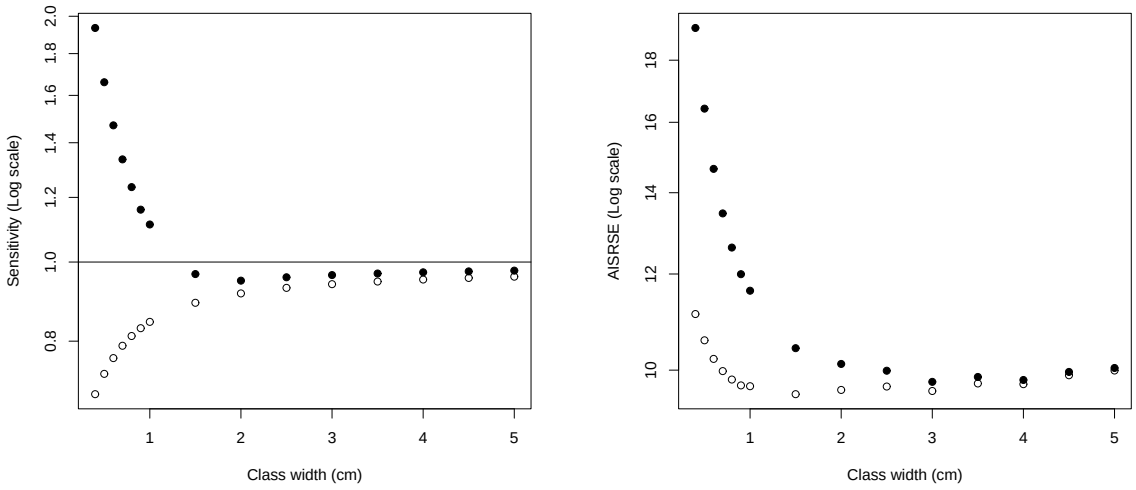
Usher's hypothesis puts a constraint on the choice of classes. At Paracou, Favrichon (1998) recommended to group individuals into 11 diameter classes ranging from 10 cm to 60 cm, with a constant amplitude equal to 5 cm. The last diameter class groups all the trees with a diameter greater than 60 cm. This choice lies on a practical justification and, with a time step equal to two years, it insures that Usher's hypothesis is realistic (0.4% of the data do not comply with Usher's hypothesis with such settings). We now intend to relax Usher's hypothesis by using the untrimmed estimator. We thus compared the properties of the classical and of the untrimmed estimators depending of the class width.

We considered diameter classes with a constant amplitude, the last class being the same as Favrichon's. The amplitude was varied from 0.4 cm to 5 cm. For the classical estimator, the upper boundary M of trimming for a given class is the upper bound of the following class, and the lower boundary M' is the lower bound l_0 of the class. This is equivalent with trimming individuals that move backwards or that move up by more than one class; in other words, it ensures that Usher's hypothesis is met. As individuals of the last two classes cannot move more than one class and as the last class is not limited, M is equal to $+\infty$ for the last two classes.

We compared the robustness and the asymptotic standardized root square error of the two estimators of transition parameters depending on the class width. More precisely, for a given class amplitude and for a given estimator, we calculated the sensitivity of the estimator in each class. Then, we calculated the maximum of these sensitivities. The asymptotic standardized root square error for a given class and a given estimator is defined as the ratio of the square root of the asymptotic variance of the estimator over its expectation.Similarly, the asymptotic standardized root square errors for each class were summed on the classes, which gave the asymptotic integrated standardized root square error (AISRSE).

5.3 Results

Results are given in Figure 2. The first graphic shows the maximum of the gross error sensitivities and the second shows the AISRSE of both estimators as a function of the class width. Firstly, the maximum gross error sensitivity of the untrimmed estimator is smaller than that of the trimmed estimator, which follows from Proposition 2. It grows until to stabilize close to one for a class width greater than 3 cm. Considering Corollary 1, this behavior can



(a) (b)

Fig. 2. Maximum of the gross error sensitivities (a) and AISRSE (b) as a function of class width. Black dots shows the trimmed estimator, whereas empty dots the untrimmed estimator.

be understood as the wider the classes are, the smaller $T(H)$ is, and then the closer to one $1 - T(H)$ gets. The maximum gross error sensitivity of the trimmed estimator decreases until 2 cm in class width and then increases slowly. For class widths greater than 3 cm, it gets close to the maximum gross error sensitivity of the untrimmed estimator. Furthermore, the AISRSE of the untrimmed estimator is smaller than that of the trimmed estimator. Both decrease until a class width of 1.5 cm, and then oscillated around the value 10. Their behavior are very similar for a class width greater than 3 cm.

In view of these results, we deduce that the untrimmed estimator is not only more robust but also smaller in quadratic error than the trimmed estimator at Paracou. Nevertheless, for a class width greater than 3 cm their behavior in sensitivity and in quadratic error is similar. This comes from the fact that the wider the class is, the smaller the proportion of individuals that violates Usher's hypothesis is. In conclusion, for the example of Paracou, the data treatment that consists in rectifying data that violate Usher's hypothesis gives better estimates in terms of robustness and quadratic error than the data treatment that consists in removing these data. This is particularly true for class widths smaller than 3 cm.

6 Discussion

In this paper, we have compared two ways of treating data that violate Usher's hypothesis. We have proved that rectifying these data gives more robust estimators of transition parameters than removing them. Nevertheless, the quadratic error of the untrimmed estimator can be greater than that of the classical estimator, and its expectation can deviate from that of the classical estimator. Furthermore, as the proportion of individuals that violate Usher's hypothesis decreases as the class width grows, the classical truncated estimator becomes similar to the untrimmed estimator for class widths great enough. Then, the untrimmed estimator is, in general, more favorable for small class widths.

At the opposite, the smaller the class width is, the greater the proportion of individuals that violate Usher's hypothesis is. Then the choice of a too small class width can lead to a too great proportion, and return a model too far from reality.

The question of the choice of the class width is also left for future work. A criterion that would allow to determine the optimal choice has to be found. Vandermeer (1978) and Moloney (1986) have proposed a criterion based on the minimization of the quadratic error of estimators. Nevertheless, this criterion is not satisfactory for the model at Paracou (Favrichon, 1995). Indeed, it gives an optimal class width equal to zero. In terms of robustness, the study of the sensitivities of transition parameters at Paracou showed that a class width equal to 2 cm was optimal for the trimmed estimator (Figure 2). In terms of quadratic error, the greater the class width is, the smaller the error is, at Paracou. Nevertheless, for a given quadratic error, using the untrimmed estimator permits to reduce the class width with respect to the classical trimmed estimator.

References

Favrichon, V., 1995. Modèle matriciel déterministe en temps discret. Application à l'étude de la dynamique d'un peuplement forestier tropical humide (Guyane française). Thèse de doctorat, Université Claude Bernard - Lyon I, Lyon, France.

Favrichon, V., 1998. Modeling the dynamics and species composition of tropical mixed-species uneven-aged natural forest: Effects of alternative cutting regimes. For. Sci. 44(1), 113–124.

Gourlet-Fleury, S., Guehl, J.M., Laroussinie, O. (Eds.), 2004. Ecology and Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana. Elsevier, Paris.

Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Am. Statist. Assoc. 69, 383–393.

Lefkovitch, L.P., 1965. The study of population growth in organisms grouped by stages. Biometrics 21, 1–18.

Leslie, P.H., 1945. In the use of matrices in certain population mathematics. Biometrika 33, 183–212.

Moloney, K.A., 1986. A generalized algorithm for determining category size. Œcologia 69, 176–180.

Usher, M.B., 1966. A matrix approach to the management of renewable resources, with special reference to the selection forests. J. Appl. Ecol. 3, 355–367.

Usher, M.B., 1969. A matrix model for forest management. Biometrics 25(2), 309–315.

Vandermeer, J., 1978. Choosing category size in a stage projection matrix. Œcologia 32, 79–84.

Annexe D

Article soumis à *Acta Biotheoretica*

Asymptotic distribution of density-dependent stage-grouped population dynamics models

Mélanie Zetlaoui[a], Nicolas Picard[b], Avner Bar-Hen[a,c,*]

[a] Institut National Agronomique Paris-Grignon, 16 rue Claude Bernard, 75231 Paris Cedex 5, France

[b] Cirad, Campus de Baillarguet, TA 10/D, 34398 Montpellier Cedex 5, France

[c] Université Paris 13, SMBH-Lim & Bio, 74 rue Marcel Cachin, 93017 Bobigny Cedex, France

∗ Corresponding author: INA-PG, OMIP, 16, rue Claude Bernard, 75231 Paris Cedex 5. Tel: (+33) 144 08 18 88. E-mail: avner.bar-hen@smbh.univ-paris13.fr

Abstract

Matrix models are widely used in biology to predict the temporal evolution of stage-structured populations. One issue related to matrix models that is often disregarded is the sampling variability. As the sample used to estimate the vital rates of the models are of finite size, a sampling error is attached to parameter estimation, which has in turn repercussions on all the predictions of the model. In this study, we address the question of building confidence bounds around the predictions of matrix models due to sampling variability. We focus on a density-dependent Usher model, the maximum likelihood estimator of parameters, and the predicted stationary stage vector. The asymptotic distribution of the stationary stage vector is specified, assuming that the parameters of the model remain in a set of the parameter space where the model admits one unique equilibrium point. Tests for density-dependence are also incidentally provided. The model is applied to a tropical rain forest in French Guiana.

Keywords: asymptotic distribution, density-dependence, matrix model, sampling variability, Usher model.

1 Introduction

Models of population dynamics are useful to quantify the evolution of population traits and predict their future. When the population is structured in stages, its evolution can be described as the evolution of the distribution of the individuals between stages. When stages and time are discrete, the corresponding dynamical system can be written as a matrix relationship (van Groenendael et al., 1988; Caswell, 2001). Transition rates between stages are the elements of the matrix. Matrix models have been widely used in ecology to deal with invasive species (Neubert and Caswell, 2000), population viability (Boyce, 1992; Alvarez-Buylla et al., 1996; Menges, 2000; Morris and Doak, 2002), or the management of harvested populations (Doubleday, 1975; Buongiorno and Gilless, 2003; Hauser et al., 2006). The most general case of matrix models corresponds to the Lefkovitch (1965) model, where any transition between stages is allowed. The transition matrix then is full. The Usher (1966, 1969) model relies on Usher's hypothesis that restricts the possible transitions: during one time step, an individual either stays alive in the same stage, moves up to the next stage, or dies. The transition matrix then has non-null elements only on its diagonal and its sub-diagonal. Usher's hypothesis is realistic when stages are ordered size classes and when growth is a first-order process. It is rarely consistent with ontogenic stages. Usher model has attracted much attention in forestry where Usher's hypothesis is consistent with tree diameter classes. Foresters have intensely used it to assess the impact of management rules (Michie and Buongiorno, 1984; Vanclay, 1995; Buongiorno et al., 1995; Boscolo and Vincent, 1998; Buongiorno and Gilless, 2003), do economic studies (Lu and Buongiorno, 1993; Buongiorno et al., 1994) or predict diversity changes in forests (Ingram and Buongiorno, 1996; Lin et al., 1996; Liang et al., 2005). The Leslie (1945) model is, from a mathematical point of view, a special case of the Usher model. In this third category of matrix models, the only possible transitions are from one stage to the next one, or death. The transition matrix then has non-null elements only on its sub-diagonal. The Leslie model corresponds to situations where stages are age classes, and has intensely been used in human demography (Keyfitz, 1967) and animal population dynamics.

Matrix models are basically defined in a deterministic way. However randomness can be introduced to enlarge their range of applications. Three types of randomness are generally distinguished (Alvarez-Buylla et al., 1996), corresponding to three different goals: demographic randomness, environmental randomness, and sampling randomness. Demographic randomness is used to deal with small-sized populations, to model for instance their probability of extinction (Boyce, 1992; Lande, 1993; Menges, 2000). It consists in

simulating the change of a population with n individuals as the sum of n independent Markov chains.

Environmental randomness consists in randomly drawing the parameters of the matrix model according to a law that gives account of environmental variations. It has often been taken into account in population viability analyses (Boyce, 1992; Fieberg and Ellner, 2001; Kaye and Pyke, 2003; Ramula and Lehtilä, 2005). It can be simulated as random shocks (Jiang and Shao, 2004; Zhou and Buongiorno, 2004; Rollin et al., 2005), as matrix sampling, or by adjusting a specific law for each parameter (Fieberg and Ellner, 2001; Kaye and Pyke, 2003; Ramula and Lehtilä, 2005).

Sampling randomness issue from the fact that the parameters of the matrix model are estimated from a sample of observations of finite size. They are thus estimators, that is to say random variables whose distribution depends on the distribution of the observations. Then, any prediction of the matrix model, that is a function of the parameters, is also a random variable. Sampling randomness is almost always disregarded (Alvarez-Buylla et al., 1996), although it can have significant effects (Taylor, 1995). It can be simply simulated by drawing samples of data and estimating, for each sample, the parameters of the matrix model. Then, conditionally to a set of values of the parameters, the matrix model is run in a deterministic way. Sampling randomness is important in applications as it relates the quantity of data available to the precision of the prediction of the quantity of interest. It permits to answer questions such as: How many individuals do we have to monitor to predict the stationary stage distribution of the population with a 10% precision at level 5%? Or: Given the sample size, what are the confidence bounds around the asymptotic growth rate of the population?

Sampling randomness thus relates to the size of the sample used to estimate vital rates, whereas demographic randomness relates to the size of the simulated population. The smaller the sample size is, the greater sampling randomness is. The smaller the simulated population is, the greater demographic randomness is. Some care should thus be taken in situations where the simulated population is as the same time the population monitored to get data, and when the size of this population is small. This is frequent when studying endangered species. In this case, it not clear which one of sampling randomness or of demographic randomness prevails.

As demographic randomness has intensely been studied whereas sampling randomness is most often disregarded, this study aims at determining the variance of model predictions due to sampling randomness. Taylor's expansion has been used to approximate the bias and variance of predictions in matrix models (Houllier et al., 1989; Alvarez-Buylla and Slatkin, 1991, 1993, 1994). Furthermore, the delta method can be used to get the exact

asymptotic law of the model predictions (Fieberg and Ellner, 2001; Zetlaoui et al., 2006). So far, these results have been obtained for linear matrix models. In this study, we extend them to density-dependent matrix models that are non-linear models whose transition rates depend on the current state of the population. We shall specify the asymptotic law of the stationary stage distribution. Incidently, a test for density-dependence is also provided. The theoretical results are applied to a tropical rain forest in French Guiana, to build confidence bounds around the predicted stationary diameter distribution.

2 Methods

2.1 The density-dependent Usher model

Size is categorized into m ordered stages. The population is described by the vector $\mathbf{N}_t = (N_{1t}, ..., N_{mt})$, where N_{it} is the number of individuals in stage i at time t. Time is discrete. Usher's hypothesis states that between time t and $t+1$, an individual either (i) stays in the same stage, (ii) grows up to the next stage, or (iii) dies. Moving backwards or growing up by more than one stage is not allowed. The deterministic Usher model then is (Usher, 1966, 1969; Caswell, 2001):

$$\mathbf{N}_{t+1} = \mathbf{U}_t \, \mathbf{N}_t$$

The Usher matrix $\mathbf{U}$ has non null elements on its diagonal, on its sub-diagonal and on its first row:

$$\mathbf{U} = \begin{pmatrix} p_1 + f_1 & f_2 & \cdots & f_m \\ q_1 & p_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & q_{m-1} & p_m \end{pmatrix}$$

where p_i is the probability for an individual to stay alive in stage i, q_i is the probability to move up from stage i to $i+1$, and f_i is the fecundity of stage i. The probability of dying for an individual in stage i is $1 - p_i - q_i$. In some situations it is not possible to estimate a fecundity for each stage, but only an average fecundity: $f = f_1 = ... = f_m$. In what follows we shall use an average fecundity f for sake of simplicity, but the results that we shall present can be straightforwardly extended to stage-specific fecundities.

Density-dependence means that the Usher matrix at time t (or, equivalently, some of the vital rates p_i, q_i and f) depends on the vector $\mathbf{N}_t$ of the number of individuals by stage:

$$\mathbf{U}_t \equiv \mathbf{U}(\mathbf{N}_t)$$

Density-dependence can take an infinity of shapes, but the most commonly found (Caswell, 2001) are the Berverton-Holt dependence:

$$v_t = \frac{c}{1 + b\,\mathbf{N}_t.\mathbf{a}}$$

and the Ricker (or exponential) dependence:

$$v_t = c\,\exp(-b\,\mathbf{N}_t.\mathbf{a}) \tag{1}$$

where v is any of the vital rates, c and b are constant parameters, $\mathbf{a}$ is a constant positive vector of $\mathbf{R}^m$, and dot denotes the scalar product in $\mathbf{R}^m$. For instance, $\mathbf{a}$ can be the unit vector $(1, 1, ..., 1)$, so that $\mathbf{N}_t.\mathbf{a}$ is the total number of individuals at time t. In forestry applications, $\mathbf{a}$ is often taken as the vector whose ith element is the average basal area of an individual in stage i. The quantity $\mathbf{N}_t.\mathbf{a}$ then represents the cumulative basal area of the population. Exploratory analyses showed us that the exponential density-dependence better suited the experimental data at Paracou than the Beverton-Holt dependence. In what follows, we shall thus focus on the exponential dependence (1).

The quantity $\mathbf{N}_t.\mathbf{a}$ is sometimes called the competition index since, if $b > 0$, the derivative of v_t with respect to this quantity is negative. An important point is that $\mathbf{N}_t.\mathbf{a}$ is cumulative and quantifies the stocking of the population as a whole. It thus implicitly depends on the spatial extension of the domain where the population lives. All things being equal, the value of $\mathbf{N}_t.\mathbf{a}$ on a 10 ha plot will be ten times that of $\mathbf{N}_t.\mathbf{a}$ on a 1 ha plot. Thus, the definition of a density-dependent matrix model must also specify the area of the plot where the population lives.

The most general model is obtained when each of the vital rates has its own density-dependence, that is to say:

$$\begin{aligned} p_{it} &= \mu_i\,\exp(-\xi_i\,\mathbf{N}_t.\mathbf{a}) \quad (i = 1...m) \\ q_{it} &= \nu_i\,\exp(-\kappa_i\,\mathbf{N}_t.\mathbf{a}) \quad (i = 1...m-1) \\ f_t &= \alpha\,\exp(-\beta\,\mathbf{N}_t.\mathbf{a}) \end{aligned} \tag{2}$$

We shall refer to this model (2) as the complete model. It has $4m$ parameters. More parsimonious models can be obtained by identifying some of the parameters μ_i, ξ_i, ν_i or κ_i. In particular, previous studies (Favrichon, 1998) showed that an adequate model for Paracou was obtained when all the ξ_i are equal, all the κ_i are equal, and μ_i and ν_i have an second-order polynomial dependence on i:

$$\begin{aligned} p_{it} &= (\mu_0 + \mu_1 i + \mu_2 i^2)\,\exp(-\xi\,\mathbf{N}_t.\mathbf{a}) \quad (i = 1...m) \\ q_{it} &= (\nu_0 + \nu_1 i + \nu_2 i^2)\,\exp(-\kappa\,\mathbf{N}_t.\mathbf{a}) \quad (i = 1...m-1) \\ f_t &= \alpha\,\exp(-\beta\,\mathbf{N}_t.\mathbf{a}) \end{aligned} \tag{3}$$

This latter model (3) will be referred as the simplified model. It has 10 parameters.

2.2 Model predictions

The density-dependent Usher models present a great variety of possible asymptotic dynamics: equilibrium points, cycles, pseudo-cycles or chaos (Caswell, 2001, chapter 16). In this study, we make the assumption that the parameter values remain in a domain of the parameter space such that, for every parameter values, there is one unique equilibrium point of the model. This hypothesis will be empirically checked by drawing the bifurcation diagram of the model. Under this simplifying hypothesis, $\mathbf{N}_t$ tends, as time tends to infinity, towards $\mathbf{N}_\infty$ such that:

$$\mathbf{N}_\infty = \mathbf{U}(\mathbf{N}_\infty)\,\mathbf{N}_\infty$$

Hence, in the stationary state $\mathbf{N}_\infty$ of the model, the dominant eigenvalue of the Usher matrix $\mathbf{U}(\mathbf{N}_\infty)$ is one, and $\mathbf{N}_\infty$ is an eigenvector corresponding to this eigenvalue. The model prediction on which we shall focus in this study is $\mathbf{N}_\infty$.

Let us now clarify the relationship between $\mathbf{N}_\infty$ and the parameters of the model. Let θ denote in short the vector of the parameters of the model, and n_θ the length of θ. To emphasize that the Usher matrix depends on θ, we shall denote $\mathbf{U}(\mathbf{N};\theta)$ as an equivalent of $\mathbf{U}(\mathbf{N})$. Let Φ denote the function from $\mathbf{R}^m \times \mathbf{R}^{n_\theta}$ into $\mathbf{R}^m$ such that

$$\Phi(\mathbf{N},\theta) = \mathbf{N} - \mathbf{U}(\mathbf{N};\theta)\,\mathbf{N} \tag{4}$$

Then the stationary stage distribution $\mathbf{N}_\infty$ is defined as an implicit function of θ by $\Phi(\mathbf{N}_\infty,\theta) = \mathbf{0}$.

2.3 Statistical model

As the parameters θ are estimated from a sample of observations, model's predictions actually rely on an estimator $\hat{\theta}$ of θ. This estimator is a random vector whose distribution depends on the distribution of the observations. The stochastic Usher matrix model is obtained by replacing the parameters θ in the expression of $\mathbf{U}$ by their estimator $\hat{\theta}$. This stochasticity corresponds to sampling stochasticity. Then, any prediction of the Usher model is also a random variable.

To complete the description of the model, the distribution of observations has to be specified. In density-independent matrix models, an observation

corresponds to the states of an individual at two consecutive time steps. In density-dependent models, it has to be completed by the stocking $\mathbf{N}.\mathbf{a}$ of the plot to which the individual belongs. Various data structures can be encountered depending on whether the observations originate from common plots or from distinct plots, and whether the same plots are found at several time steps or not. If the same plot is followed along time, then the data are longitudinal. For sake of simplicity, we restrict this study to non-longitudinal data. The data thus originate from K distinct plots that are supposed independent. Each plot k contains n_k independent observations and the sample size is $n = \sum_{k=1}^{K} n_k$.

The distribution of the observations can be seen as a mixture model, where the mixture proportions are the proportions of the number of observations in each plot. For sake of simplicity again, we shall consider that these proportions are fixed. A slightly more complex alternative would be to specify a distribution for the number of individuals in each plot.

Given these assumptions, an observation can be defined as a random triple (i, j, k) where k is the plot to which the individual belongs, i is the stage of the individual at initial time, and j is its stage one time step later. Death will be denoted $\dagger$. A newly recruited individual in plot k will be denoted $(0, 1, k)$. The set of all possible events is $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, where $\mathcal{A}_1 = \{(i, i, k), 1 \leq i \leq m, 1 \leq k \leq K\}$ gathers the events "the individual stays in the same stage", $\mathcal{A}_2 = \{(i, i+1, k), 1 \leq i \leq m-1, 1 \leq k \leq K\}$ gathers the events "the individual moves up to the next stage", $\mathcal{A}_3 = \{(i, \dagger, k), 1 \leq i \leq m, 1 \leq k \leq K\}$ gathers the events "the individual dies", and $\mathcal{A}_4 = \{(0, 1, k), 1 \leq k \leq K\}$ gathers the events "the individual is recruited in the first stage". The number of distinct events is $3mK$. An observation, denoted X, thus is a random triple taking value in $\mathcal{A}$. Conditionally to its belonging to plot k, its distribution is:

$$\begin{array}{rcl} \Pr[X = (i, i, k)] & = & (1 - f_k^\star)\, p_{ik}\, d_{ik} \\ \Pr[X = (i, i+1, k)] & = & (1 - f_k^\star)\, q_{ik}\, d_{ik} \\ \Pr[X = (i, \dagger, k)] & = & (1 - f_k^\star)\, (1 - p_{ik} - q_{ik})\, d_{ik} \\ \Pr[X = (0, 1, k)] & = & f_k^\star \end{array} \tag{5}$$

where d_{ik} is the proportion of individuals of plot k in stage i ($\sum_{i=1}^{m} d_{ik} = 1$, $\forall k$), p_{ik} is the proportion of individuals of stage i of plot k staying in stage i, q_{ik} is the proportion of individuals of stage i of plot k growing from stage i to $i+1$, and $f_k^\star$ is the proportion of individuals of the sub-sample of plot k that have been recruited. We also use the convention $q_{mk} = 0$. Given the stocking $\mathbf{N}_k.\mathbf{a}$ of plot k, p_{ik} and q_{ik} are obtained by replacing $\mathbf{N}_t.\mathbf{a}$ by $\mathbf{N}_k.\mathbf{a}$ in the expressions (2) (for the complete model) or (3) (for the simplified model) of p_{it} and q_{it}. Let f_k be the average fecundity of plot k obtained from

(2) or (3). The difference between f_k and $f_k^\star$ is that f_k is the ratio of the number of recruited trees over the total number of individuals (excluding the recruited trees themselves) whereas $f_k^\star$ is the ratio of the same number over the sub-sample size for plot k (thus including the recruited trees themselves):

$$\frac{f_k}{f_k^\star} = \frac{n_k}{n_k - n_k f_k^\star} \Rightarrow f_k^\star = \frac{f_k}{1+f_k} \tag{6}$$

A sample of $n = \sum_{k=1}^K n_k$ observations is obtained as the union of K independent sub-samples, where the kth sub-sample is composed of n_k independent observations identically distributed according to (5).

2.4 Parameter estimation

Parameters θ are estimated using their maximum-likelihood (ML) estimator (Gross et al., 2006). Given a sample X_1, ..., X_n of observations, the model's likelihood L_θ verifies $\ln L_\theta(X_1, ..., X_n) = \sum_{j=1}^n \ln \ell_\theta(X_j)$, where

$$\begin{aligned} \ell_\theta(x) \;=\; & \sum_{k=1}^K \rho_k \left\{ \sum_{i=1}^m (1-f_k^\star) p_{ik} d_{ik} \mathbf{1}_{x=(i,i,k)} + (1-f_k^\star) q_{ik} d_{ik} \mathbf{1}_{x=(i,i+1,k)} \right. \\ & \left. + (1-f_k^\star)(1-p_{ik}-q_{ik}) d_{ik} \mathbf{1}_{x=(i,\dagger,k)} + f_k^\star \mathbf{1}_{x=(0,1,k)} \right\} \end{aligned} \tag{7}$$

ρ_k is the fixed proportion of observations in plot k, and by convention $q_{mk} = 0$. The ML estimator of θ thus is:

$$\begin{aligned} \hat{\theta} \;=\; & \arg\max_\theta \left\{ \sum_{k=1}^K \left[R_k \ln f_k^\star + N_k \ln(1-f_k^\star) \right] \right. \\ & \left. + \sum_{i=1}^m \sum_{k=1}^K \left[F_{iik} \ln p_{ik} + F_{ii+1k} \ln q_{ik} + F_{i\dagger k} \ln(1-p_{ik}-q_{ik}) \right] \right\} \end{aligned} \tag{8}$$

where $R_k = \sum_{j=1}^n \mathbf{1}_{X_j=(0,1,k)}$ is the number of recruited individuals in plot k, $N_k = n_k - R_k$, $F_{iik} = \sum_{j=1}^n \mathbf{1}_{X_j=(i,i,k)}$ is the number of individuals that stay in stage i in plot k, $F_{ii+1k} = \sum_{j=1}^n \mathbf{1}_{X_j=(i,i+1,k)}$ is the number of individuals that move up from stage i to $i+1$ in plot k, and $F_{i\dagger k} = \sum_{j=1}^n \mathbf{1}_{X_j=(i,\dagger,k)}$ is the number of individuals that die in stage i in plot k.

In the right hand side of (8), only the first sum depends on the parameters α et β. The ML estimators of α and β thus maximize

$$\sum_{k=1}^K R_k \ln f_k^\star + N_k \ln(1-f_k^\star)$$

with respect to α and β. Similarly, in the complete model (2), the parameters of each stage can be estimated separately from the other stages: $\forall i$, the ML estimators of μ_i, ξ_i, ν_i and κ_i maximize

$$\sum_{k=1}^{K} F_{iik} \ln p_{ik} + F_{ii+1k} \ln q_{ik} + F_{i\dagger k} \ln(1 - p_{ik} - q_{ik})$$

with respect to μ_i, ξ_i, ν_i and κ_i. This is no longer possible in the simplified model (3), since all the stages depend on common parameters a_i and b_i ($i = 0$, 1, 2).

The maximization of the likelihood is done numerically using a Nelder-Mead algorithm (Nelder and Mead, 1965). The advantage of the complete model (2) is that numerical optimization is done with respect to four parameters at the most: the $4m$ parameters are estimated using $m-1$ optimization with respect to four parameters and two optimization with respect to two parameters. On the contrary the simplified model (3) requires one optimization with respect to eight parameters simultaneously, and one optimization with respect to two parameters. The trapping of the optimization algorithm in local maxima is thus more likely to occur for the simplified model than for the complete model.

On the other hand, the complete model requires that observations be present for every stage. If no observation is available for one stage, then the parameters for this stage cannot be estimated. On the contrary, the simplified model can handle situations where observations are lacking for some stages. Its parameters are then still estimable.

2.5 Testing density-dependence and model selection

Density-dependence can be tested by testing whether ξ_i, κ_i, β (complete model) or ξ, κ, β (simplified model) are null. As this implies nested models, it can be tested using likelihood ratio tests. Thus, $2m$ null hypotheses were tested for the complete model: $\xi_i = 0$ ($1 \leq i \leq m$), $\kappa_i = 0$ ($1 \leq i \leq m-1$), and $\beta = 0$. For the simplified, we tested $\xi = 0$, $\kappa = 0$ and $\beta = 0$. In addition, $\mu_2 = 0$ and $\nu_2 = 0$ were also tested to ensure that a second-order polynomial was indeed required in (3).

Furthermore, the Bayesian information criterion (BIC) and the Akaike information criterion (AIC) were used to compare the complete model with the simplified model, and choose the most relevant one.

2.6 The Paracou data set

The density-dependent Usher model was applied to a tropical rain forest in French Guiana. Data came from the Paracou experimental site (5°15′N, 52°55′W) that consists of 48 permanent plots of size 125 m × 125 m. The plots were created in 1984. All trees with a diameter at breast height greater than 10 cm, including ingrowth, have been monitored. The spatial coordinates and the species of a tree were recorded when it was recruited. Its diameter has been measured annually (every two years since 1995), unless it died. In 1987, 36 plots underwent silvicultural treatments with three levels of logging intensity (12 repetitions for each level). The 12 remaining plots were left as control. More precisions on the Paracou experimental site can be found in Gourlet-Fleury et al. (2004). In this study, we used the data of 1988 and 1990, just after logging, to get contrasted situations in terms of plot stocking and population dynamics.

Favrichon (1998) has developed a density-dependent Usher model for Paracou and we thus followed his recommandations. The spatial extension for a population is a 125 m × 125 m plot. The plot stocking is computed using $\mathbf{a} = (1, ..., 1)$, so that $\mathbf{N}.\mathbf{a}$ is the total number of trees on the plot. The time step is two year. Individuals are distributed along 11 diameter classes ranging from 10 cm to 60 cm, with a constant class width of 5 cm except the last class that groups all the trees with a diameter greater than 60 cm. This time step and these breakpoints of the diameter classes ensure that Usher's hypothesis is verified.

The data set consists of 39,799 trees. For each tree, the diameter in 1988 and 1990 is given, unless the tree was dead in 1990 or not yet recruited in 1988. In the former case, only the diameter in 1988 is given, and in the latter case, only the diameter in 1990 is given. A few individuals (240 trees, that is 0.6%) did not comply with Usher's hypothesis. Following Zetlaoui et al. (in press), these data were transformed as follows: trees that moved backwards were considered as having stayed in the same class, and trees that grew up by more than one class were considered as having grown by one class. As the data did not allow to link a recruited tree to its parent, an average fecundity was estimated.

3 Results

3.1 Asymptotic distribution of model predictions

The model prediction on which we focus is the stationary stage distribution $\mathbf{N}_\infty$, but the reasoning could be extended to any other model prediction.

Assuming that it exists, $\mathbf{N}_\infty$ is an implicit function $h(\theta)$ of the parameters θ of the model, such that $\Phi[h(\theta), \theta] = 0$ where Φ is given by (4). Then, if $\hat{\theta}_n$ is the ML estimator of θ for a sample of n observations, by the functional invariance of ML estimates, $\hat{\mathbf{N}}_{\infty n} = h(\hat{\theta}_n)$ is the ML estimator of $\mathbf{N}_\infty$. The asymptotic behaviour of $\hat{\mathbf{N}}_{\infty n}$ is then given by the following proposition.

Proposition 1 *With the notations above,*

- *assuming that there exists a neighbourhood ϑ_θ of θ in the parameter space such that, $\forall \theta' \in \vartheta_\theta$, the density-dependent Usher model with parameters θ' admits one unique equilibrium point with stage distribution $h(\theta')$,*

- *assuming that, given the distribution of observations, the set of possible values of $\hat{\theta}$ is included in ϑ_θ,*

then $\sqrt{n}(\hat{\mathbf{N}}_{\infty n} - \mathbf{N}_\infty)$ converges in law towards a multinormal law in $\mathbf{R}^m$ with mean zero and asymptotic covariance matrix:

$$\mathbf{S}^2 = (\mathrm{d}_\theta h) \cdot I(\theta)^{-1} \cdot {}^{\mathrm{t}}(\mathrm{d}_\theta h)$$

where ${}^{\mathrm{t}}\mathbf{M}$ is the transpose of the matrix $\mathbf{M}$, $I(\theta)$ is the Fisher information matrix of θ for a sample of size one, $\mathrm{d}_\theta h$ is the differential of h at θ represented by a $m \times n_\theta$ matrix, and dot indicates the matrix multiplication.

Proof. As $\hat{\theta}_n$ is the ML estimator of θ for a sample of n observations, $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in law as n tends to infinity towards a multinormal law with mean zero and asymptotic covariance matrix $I(\theta)^{-1}$. The assumptions of the proposition ensure that the mapping h of θ onto $\mathbf{N}_\infty$ is an application. Then, as Φ is differentiable, the implicit function theorem proves that h is differentiable. As $\hat{\mathbf{N}}_{\infty n} = h(\hat{\theta}_n)$, the proposition is finally obtained from the asymptotic behaviour of $\hat{\theta}_n$ using the delta method. ■

The expression of $\mathrm{d}_\theta h$ and $I(\theta)$ for the simplified model (3) are given in Appendix A and B. Anticipating on the subsequent results, the simplified model turned to be more relevant at Paracou than the complete model. This is why, to conserve space, we do not give here the expression of $\mathrm{d}_\theta h$ and $I(\theta)$ for the complete model (but they can be computed in a similar way).

3.2 Application to Paracou

The estimates of parameters related to transition rates for the complete model at Paracou, together with the results of the likelihood ratio tests,

are given in Table 1. For the simplified model, the estimates of parameters related to transition rates are given in Table 2. For both models, the estimates of fecundity parameters are: $\hat{\alpha} = 0.8864$ and $\hat{\beta} = 4.0011 \times 10^{-3}$.

Tables 1 and 2 to be inserted about here

The tests of the density-dependence at Paracou reject on the whole density-independence for both the complete and the simplified model. At level 5%, the 22 hypotheses ($\xi_i = 0, 1 \leq i \leq m, \kappa_i = 0, 1 \leq i \leq m-1$, and $\beta = 0$) are all rejected except $\xi_{10} = 0$, $\kappa_9 = 0$, and $\kappa_{10} = 0$ (Table 1). As the 22 tests are not mutually independent, they cannot be interpreted separately. Nevertheless, this result suggests that the impact of competition on tree dynamics is weaker in the greater diameter classes. For the simplified model, the null hypotheses ($\xi = 0, \kappa = 0$, and $\beta = 0$) are all rejected (Table 2). As the parameters are the same for all classes, the hypothesis of density-dependence is accepted in this model for all classes simultaneously. In addition, the hypotheses $\mu_2 = 0$ and $\nu_2 = 0$ were rejected (Table 2), which shows that a second-order polynomial was indeed required in (3).

The complete and the simplified models were compared using the Akaike and the Bayesian information critera. According to the AIC, the complete model (AIC = 31975.18) is better than the simplified model (AIC = 32023.54). The opposite conclusion is obtained with the BIC (BIC=32334.44 for the complete model and BIC=32091.97 for the simplified model). These results are consistent with the tendency of AIC to favour over-parameterization while BIC tends to penalize over-parameterization (Liquet, 2002). However the differences are small in both cases, so that the two models can be considered as much relevant. As the simplified model has less parameters than the complete model and following Occam's razor, we hereafter focus on the simplified model.

The estimate of N_∞ predicted by the simplified model at Paracou is $\hat{N}_\infty$ = (279.0, 170.3, 106.0, 66.8, 42.5, 27.2, 17.5, 11.2, 7.2, 4.6, 2.9), where the numbers of trees are given for a plot 125 m × 125 m (1.5625 ha). The covariance matrix of $\hat{N}_\infty$ based on the asymptotic expression given in Proposition 1 and sample size $n = 39,799$ is given in Table 3. In the present case, the sample size is so large that the asymptotic approximation can be reasonably used. The correlation between two classes increases as the two classes get closer. Moreover the correlation between two neighbour classes increases from smaller to larger diameter classes.

Table 3 to be inserted about here

As the numbers $\hat{N}_{\infty i}$ of trees in each class are mutually dependent, one should consider a confidence ellipsoid in $\mathbf{R}^{11}$ around $\hat{\mathbf{N}}_\infty$ rather than separate confidence intervals for each $\hat{N}_{\infty i}$. Moreover, as $\hat{\mathbf{N}}_\infty$ is approximately multinormal, any linear combination of its elements is approximately normal. For instance the total density $\hat{N}_{\text{tot}} = \hat{\mathbf{N}}_\infty.\mathbf{1}$, where $\mathbf{1}$ is the vector of length 11 whose elements are equal to one, is approximately normal with mean $\mathbf{N}_\infty.\mathbf{1}$ and variance ${}^{\text{t}}\mathbf{1} \cdot \mathbf{S}^2 \cdot \mathbf{1}$. Similarly, the cumulative basal area $\hat{B}_{\text{tot}} = \hat{\mathbf{N}}_\infty.\mathbf{B}$, where $\mathbf{B}$ is the vector of length 11 whose ith element is equal to $\frac{\pi}{4} D_i^2$ (where D_i is the average diameter of class i), is approximately normal with mean $\mathbf{N}_\infty.\mathbf{B}$ and variance ${}^{\text{t}}\mathbf{B} \cdot \mathbf{S}^2 \cdot \mathbf{B}$. Finally, $\hat{N}_{\text{tot}}$ and $\hat{B}_{\text{tot}}$ are correlated and their covariance equal ${}^{\text{t}}\mathbf{B} \cdot \mathbf{S}^2 \cdot \mathbf{1}$. Figure 1 shows the estimated value of $(\hat{N}_{\text{tot}}, \hat{B}_{\text{tot}})$ together with their confidence ellipse at level 5% for different sample sizes. The sample size at Paracou is 39,799. The average density in control plots is 965 trees. The sample size such that the area of the confidence ellipse at level 5% be equal to $\hat{N}_{\text{tot}}\hat{B}_{\text{tot}}/100$ is 30,154.

The observed values of $(N_{\text{tot}}, B_{\text{tot}})$ for the 48 Paracou plots in 1988 are also shown in Figure 1. The predicted stationary state is quite far from the control plots of Paracou (white circles in Fig.1). An interpretation of this result is that the plots in 1988–90 are still recovering from the 1987 logging, with many trees dying from injuries caused by logging. Then, the dynamics of the plots in 1988–90 is slower than it would be if plots were intact from logging damages.

Figure 1 to be inserted about here

4 Discussion and conclusion

Assuming that the parameters remain in a set of the parameter space where the density-dependent model has one unique equilibrium point, the asymptotic distribution of the stationary state of the density-dependent matrix model can be specified. The reasoning that we made for a specific Usher model, the stationary diameter distribution, and the ML estimator of the parameters can be easily extended to other models, other predicted quantities, and other estimators. However, the existence of a unique equilibrium point is critical. As soon as the dynamical system oscillates or is chaotic, the function $h(\theta)$ of Prop.1 is no longer differentiable with respect to the parameters θ.

Changing the model or the predicted quantity will change the function h in Prop.1. As h can be specified implicitly, as is the case here, a great flexibility is offered. Changing the estimator will change the Fisher information matrix $I(\theta)$ in Prop.1. Actually, as soon as the estimator is not the ML

estimator, its asymptotic covariance matrix is no longer given by $I(\theta)^{-1}$. If applicable, one can then use the influence curve to specify the asymptotic covariance matrix (Hampel, 1974; Cuevas and Romo, 1995).

In this study, following Gross et al. (2006), we used the ML estimator of the parameters. This provides an alternative to the classical way of estimating the parameters θ of a density-dependent Usher model. The classical procedure is a two-step procedure (Solomon et al., 1986; Favrichon, 1998): (i) the vital rates are estimated for each plot separately; let $\hat{p}_{ik}$, $\hat{q}_{ik}$ and $\hat{f}_k$ be their estimates for plot k; and (ii) the model (2) or (3) is fitted to the data $\hat{p}_{ik}$, $\hat{q}_{ik}$, $\hat{f}_k$ to get the parameters $\theta = (\mu_i, \xi_i, 1 \leq i \leq m; \nu_i, \kappa_i, 1 \leq i \leq m-1; \alpha, \beta)$ (complete model) or $\theta = (a_i, b_i, 0 \leq i \leq 2; \xi, \kappa, \alpha, \beta)$. This procedure leads to untractable estimators, whereas the ML estimator requires only one step, is mathematically tractable, and is asymptotically the most efficient. In compensation, the ML estimator requires the use of numerical optimization that can provide unstable estimates. The choice of the starting point for the optimization algorithm may in particular be critical.

In prospect, we intend to extend the results to estimators that are as the same time mathematically tractable and easy to implement. An extension of the estimator proposed by Rogers-Bennett and Rogers (2006) to the density-dependent case could be faced.

References

Alvarez-Buylla, E. R., García-Barrios, R., Lara-Moreno, C., Martínez-Ramos, M., 1996. Demographic and genetic models in conservation biology: applications and perpectives for tropical rain forest tree species. Annual Review of Ecology and Systematics 27, 387–421.

Alvarez-Buylla, E. R., Slatkin, M., 1991. Finding confidence limits on population growth rates. Trends in Ecology and Evolution 6 (7), 221–224.

Alvarez-Buylla, E. R., Slatkin, M., 1993. Finding confidence limits on population growth rates: Monte Carlo test of a simple analytic method. Oikos 68, 273–282.

Alvarez-Buylla, E. R., Slatkin, M., 1994. Finding confidence limits on population growth rates: three real examples revised. Ecology 75 (1), 255–260.

Boscolo, M., Vincent, J. R., 1998. Promoting better logging practices in tropical forests: A simulation analysis of alternative regulations. Development

Discussion Paper 652, The Harvard Institute for International Development, Harvard University, Cambridge, Massachusetts, USA.
URL http://www.hiid.harvard.edu/pub/ddps/652abs.html

Boyce, M. S., 1992. Population viability analysis. Annual Review of Ecology and Systematics 23, 481–506.

Buongiorno, J., Dahir, S., Lu, H. C., Lin, C. R., 1994. Tree size diversity and economic returns in uneven-aged forest stands. Forest Science 40 (1), 83–103.

Buongiorno, J., Gilless, J. K., 2003. Decision Methods for Forest Resource Management. Academic Press.

Buongiorno, J., Peyron, J. L., Houllier, F., Bruciamacchie, M., 1995. Growth and management of mixed-species, uneven-aged forests in the French Jura: implications for economic returns and tree diversity. Forest Science 41 (3), 397–429.

Caswell, H., 2001. Matrix Population Models: Construction, Analysis and Interpretation, 2nd Edition. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

Cuevas, A., Romo, J., 1995. On the estimation of the influence curve. The Canadian Journal of Statistics 23 (1), 1–9.

Doubleday, W. G., 1975. Harvesting in matrix population models. Biometrics 31 (1), 189–200.

Favrichon, V., 1998. Modeling the dynamics and species composition of tropical mixed-species uneven-aged natural forest: Effects of alternative cutting regimes. Forest Science 44 (1), 113–124.

Fieberg, J., Ellner, S. P., 2001. Stochastic matrix models for conservation and management: a comparative review of methods. Ecology Letters 4 (3), 244–266.

Gourlet-Fleury, S., Guehl, J. M., Laroussinie, O. (Eds.), 2004. Ecology and Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana. Elsevier, Paris.

Gross, K., Morris, W. F., Wolosin, M. S., Doak, D. F., 2006. Modeling vital rates improves estimation of population projection matrices. Population Ecology 48 (1), 79–89.

Hampel, F. R., 1974. The influence curve and its role in robust estimation. Journal of the American Statistical Association 69, 383–393.

Hauser, C. E., Cooch, E. G., Lebreton, J. D., 2006. Control of structured populations by harvest. Ecological Modelling 196 (3-4), 462–470.

Houllier, F., Lebreton, J. D., Pontier, D., 1989. Sampling properties of the asymptotic behavior of age- or stage-grouped population models. Mathematical Biosciences 95 (2), 161–177.

Ingram, D., Buongiorno, J., 1996. Income and diversity tradeoffs from management of mixed lowland dipterocarps in Malaysia. Journal of Tropical Forest Science 9 (2), 242–270.
URL http://www.fpl.fs.fed.us/econ/Publications.htm\#Ingramcd

Jiang, L., Shao, N., 2004. Red environmental noise and the appearance of delayed density dependence in age-structured populations. Proceedings of the Royal Society of London, Series B 271 (1543), 1059–1064.
URL http://marine.rutgers.edu/ebme/html_docs/reprints/Jiang%20and%20Shao%202004.pdf

Kaye, T. N., Pyke, D. A., 2003. The effect of stochastic technique on estimates of population viability from transition matrix models. Ecology 84 (6), 1464–1476.

Keyfitz, N., 1967. Reconciliation of population models: matrix, integral equation and partial fraction. Journal of the Royal Statistical Society, Series A 130, 61–83.

Lande, R., 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. The American Naturalist 142, 911–927.

Lefkovitch, L., 1965. The study of population growth in organisms grouped by stages. Biometrics 21, 1–18.

Leslie, P., 1945. In the use of matrices in certain population mathematics. Biometrika 33, 183–212.

Liang, J., Buongiorno, J., Monserud, R. A., 2005. Growth and yield of all-aged douglas-fir – western hemlock forest stands: a matrix model with stand diversity effects. Canadian Journal of Forest Research 35 (10), 2368–2381.

Lin, C. R., Buongiorno, J., Vasievich, M., 1996. A multi-species, density-dependent matrix growth model to predict tree diversity and income in northern hardwood stands. Ecological Modelling 91 (1-3), 193–211.

Liquet, B., 2002. Sélection de modèles semi-paramétriques. Thèse de doctorat, Université Victor Segalien–Bordeaux 2, Bordeaux.

Lu, H., Buongiorno, J., 1993. Long- and short-term effects of alternative cutting regimes on economic returns and ecological diversity in mixed-species forests. Forest Ecology and Management 58 (3-4), 173–192.

Menges, E. S., 2000. Population viability analyses in plants: challenges and opportunities. Trends in Ecology and Evolution 15 (2), 51–56.

Michie, B. R., Buongiorno, J., 1984. Estimation of a matrix model of forest growth from re-measured permanent plots. Forest Ecology and Management 8, 127–135.

Morris, W. F., Doak, D. F., 2002. Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis. Sinauer Associates, Inc., Sunderland, MA.

Nelder, J. A., Mead, R., 1965. A simplex algorithm for function minimization. Computer Journal 7, 308–313.

Neubert, M. G., Caswell, H., 2000. Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. Ecology 81 (6), 1613–1628.

Ramula, S., Lehtilä, K., 2005. Importance of correlations among matrix entries in stochastic models in relation to number of transition matrices. Oikos 111 (1), 9–18.

Rogers-Bennett, L., Rogers, D. W., 2006. A semi-empirical growth estimation method for matrix models of endangered species. Ecological Modelling 195 (3-4), 237–246.

Rollin, F., Buongiorno, J., Zhou, M., Peyron, J. L., 2005. Management of mixed-species, uneven-aged forests in the French Jura: from stochastic growth and price models to decision tables. Forest Science 51 (1), 64–75.

Solomon, D. S., Hosmer, R. A., Hayslett, H. T., 1986. A two-stage matrix model for predicting growth of forest stands in the Northeast. Canadian Journal of Forest Research 16, 521–528.

Taylor, B. L., 1995. The reliability of using population viability analysis for risk classification of species. Conservation Biology 9 (3), 551–558.

Usher, M. B., 1966. A matrix approach to the management of renewable resources, with special reference to the selection forests. Journal of Applied Ecology 3, 355–367.

Usher, M. B., 1969. A matrix model for forest management. Biometrics 25 (2), 309–315.

van Groenendael, J., de Kroon, H., Caswell, H., October 1988. Projection matrices in population biology. Trends in Ecology and Evolution 3 (10), 264–269.

Vanclay, J. K., 1995. Growth models for tropical forests: A synthesis of models and methods. Forest Science 41 (1), 7–42.

Zetlaoui, M., Picard, N., Bar-Hen, A., 2006. Asymptotic distribution of stage-grouped population models. Mathematical Biosciences 200 (1), 76–89.

Zetlaoui, M., Picard, N., Bar-hen, A., in press. Robustness of the estimators of transition rates for stage-classified matrix models. Computational Statistics and Data Analysis .

Zhou, M., Buongiorno, J., 2004. Nonlinearity and noise interaction in a model of forest growth. Ecological Modelling 180 (2-3), 291–304.

Appendix

A Differential of the application $\theta \mapsto \mathbf{N}_\infty$

Assuming that it is an application, h maps the vector of parameters θ onto the asymptotic stage distribution $\mathbf{N}_\infty$. We hereafter focus on the simplified model (3). Thus, $h : \mathbf{R}^{10} \to \mathbf{R}^m$, and the differential $\mathrm{d}_\theta h$ is a $m \times 10$ matrix whose element on the ith row and jth column is $\partial N_i / \partial \theta_j$, where N_i is the ith element of $\mathbf{N}_\infty$ and θ_j is the jth element of θ. The function h is implicitly defined as $\Phi[h(\theta), \theta] = 0$ where $\Phi : \mathbf{R}^m \times \mathbf{R}^{10} \to \mathbf{R}^m$ is given by (4). Let Φ_i be the ith element of Φ ($1 \leq i \leq m$), let $\mathrm{d}_\theta \Phi$ be the partial differential of Φ represented by the $m \times 10$ matrix $(\partial \Phi_i / \partial \theta_j)_{1 \leq i \leq m, 1 \leq j \leq 10}$, and let $\mathrm{d}_{\mathbf{N}} \Phi$ be the

partial differential of Φ represented by the $m \times m$ matrix $(\partial \Phi_i / \partial N_j)_{1 \leq i,j \leq m}$. Then, by the implicit function theorem,

$$\mathrm{d}_\theta h = - \left. (\mathrm{d}_{\mathbf{N}} \Phi)^{-1} \right|_{(h(\theta),\theta)} \cdot \left. \mathrm{d}_\theta \Phi \right|_{(h(\theta),\theta)}$$

We are now left with computing the partial derivatives of Φ.

For the simplified model (3) and $\forall i, j$,

$$\frac{\partial p_i}{\partial N_j} = -\xi a_j \, p_i(\mathbf{N}), \quad \frac{\partial q_i}{\partial N_j} = -\kappa a_j \, q_i(\mathbf{N}), \quad \frac{\partial f}{\partial N_j} = -\beta a_j \, f(\mathbf{N})$$

where a_j is the jth element of $\mathbf{a}$. Using the developed form of the matrix relationship,

$$\Phi_i(\mathbf{N}, \theta) = \begin{cases} N_1 - p_1(\mathbf{N}; \theta) \, N_1 - f(\mathbf{N}; \theta) \, \sum_{j=1}^m N_j & (i = 1) \\ N_i - p_i(\mathbf{N}; \theta) \, N_i - q_{i-1}(\mathbf{N}; \theta) \, N_{i-1} & (2 \leq i \leq m) \end{cases}$$

The partial derivatives of Φ with respect to $\mathbf{N}$ are thus:

- for the first element of Φ,

$$\frac{\partial \Phi_1}{\partial N_j} = \begin{cases} 1 - p_1 + \xi a_1 p_1 N_1 + \beta a_1 f \sum_{k=1}^m N_k - f & (j = 1) \\ \xi a_j p_1 N_1 + \beta a_j f \sum_{k=1}^m N_k - f & (2 \leq j \leq m) \end{cases}$$

- for $2 \leq i \leq m$,

$$\frac{\partial \Phi_i}{\partial N_j} = \begin{cases} 1 - p_i + \xi a_i p_i N_i + \kappa a_i q_{i-1} N_{i-1} & (j = i) \\ -q_{i-1} + \xi a_{i-1} p_i N_i + \kappa a_{i-1} q_{i-1} N_{i-1} & (j = i - 1) \\ \xi a_j p_i N_i + \kappa a_j q_{i-1} N_{i-1} & (j \neq i, i - 1) \end{cases}$$

Moreover, given that the parameters are in the order $\theta = (\mu_0, \, \mu_1, \, \mu_2, \, \xi, \, \nu_0, \, \nu_1, \, \nu_2, \, \kappa, \, \alpha, \, \beta)$, the partial derivatives of Φ with respect to θ are:

- for $1 \leq j \leq 4$,

$$\frac{\partial \Phi_i}{\partial \theta_j} = -\frac{\partial p_i}{\partial \theta_j} N_i \qquad (1 \leq i \leq m)$$

- for $5 \leq j \leq 8$,

$$\frac{\partial \Phi_i}{\partial \theta_j} = \begin{cases} 0 & (i = 1) \\ -(\partial q_{i-1} / \partial \theta_j) N_{i-1} & (2 \leq i \leq m) \end{cases}$$

- for $j = 9, \, 10$,

$$\frac{\partial \Phi_i}{\partial \theta_j} = \begin{cases} -(\partial f / \partial \theta_j) \sum_{k=1}^m N_k & (i = 1) \\ 0 & (2 \leq i \leq m) \end{cases}$$

where the partial derivatives of the vital rates with respect to the parameters θ are:

$$\begin{aligned}
\frac{\partial p_i}{\partial \theta_j} &= i^{j-1} \exp(-\xi \mathbf{N}.\mathbf{a}) \text{ for } 1 \leq j \leq 3, \quad \frac{\partial p_i}{\partial \theta_4} = -\mathbf{N}.\mathbf{a}\, p_i \\
\frac{\partial p_i}{\partial \theta_j} &= 0 \text{ for } 5 \leq j \leq 10 \\
\frac{\partial q_i}{\partial \theta_j} &= i^{j-5} \exp(-\kappa \mathbf{N}.\mathbf{a}) \text{ for } 5 \leq j \leq 7, \quad \frac{\partial q_i}{\partial \theta_8} = -\mathbf{N}.\mathbf{a}\, q_i \\
\frac{\partial q_i}{\partial \theta_j} &= 0 \text{ for } 1 \leq j \leq 4 \text{ or } j = 9, 10 \\
\frac{\partial f}{\partial \theta_j} &= 0 \text{ for } 1 \leq j \leq 8, \quad \frac{\partial f}{\partial \theta_9} = \frac{f}{\alpha}, \quad \frac{\partial f}{\partial \theta_{10}} = -\mathbf{N}.\mathbf{a}\, f
\end{aligned}$$

Finally, to compute the differential of h at θ it is necessary to compute $\mathbf{N}_\infty = h(\theta)$, that is to say to solve $\Phi(\mathbf{N}, \theta) = 0$ for $\mathbf{N}$. To this end, we used the numerical method proposed by Caswell (2001, § 16.2.1 p.518).

B Fisher information matrix of θ

Computations are done for the simplified model (3). The 10 parameters are taken in the order $\theta = (\mu_0,\ \mu_1,\ \mu_2,\ \xi,\ \nu_0,\ \nu_1,\ \nu_2,\ \kappa,\ \alpha,\ \beta)$. The Fisher information matrix $I(\theta)$ for a sample of size one is a 10×10 matrix with element:

$$[I(\theta)]_{ij} = \mathrm{E}\left[\left(\frac{\partial \ln \ell_\theta(X)}{\partial \theta_i}\right)\left(\frac{\partial \ln \ell_\theta(X)}{\partial \theta_j}\right)\right]$$

where ℓ_θ is given by (7) and the expectation is over the distribution (5) of the observation X. As the estimators of α and β are independent from the other parameters, $I(\theta)$ is actually of the form:

$$I(\theta) = \left(\begin{array}{c|c} I(\theta^\star) & \mathbf{0} \\ \hline \mathbf{0} & I(\alpha, \beta) \end{array}\right)$$

where $\theta^\star$ is the restriction of θ to the eight first parameters. The logarithm of ℓ_θ is:

$$\begin{aligned}
\ln \ell_\theta(X) &= \sum_{i=1}^{m}\sum_{k=1}^{K} \ln d_{ik} \mathbf{1}_{X=(i,\bullet,k)} + \sum_{k=1}^{K} [\ln(1 - f_k^\star)\mathbf{1}_{X=(\bullet,\bullet,k)} + \ln f_k^\star \mathbf{1}_{X=(0,1,k)}] \\
&+ \sum_{i=1}^{m}\sum_{k=1}^{K} [\ln p_{ik} \mathbf{1}_{X=(i,i,k)} + \ln q_{ik} \mathbf{1}_{X=(i,i+1,k)} + \ln(1 - p_{ik} - q_{ik})\mathbf{1}_{X=(i,\dagger,k)}] \\
&+ \sum_{k=1}^{K} \ln \rho_k \mathbf{1}_{X=(.,.,k)}
\end{aligned}$$

where the small dot denotes any stage (including 0), the big dot denotes any stage except 0 and, by convention, $q_{mk} = 0$. The partial derivatives of $\ln \ell_\theta$ with respect to θ then are:

$$\frac{\partial \ln \ell_\theta(X)}{\partial \theta_j} = \begin{cases} \displaystyle\sum_{i=1}^{m}\sum_{k=1}^{K} \frac{\partial p_{ik}}{\partial \theta_j}\left[\frac{\mathbf{1}_{X=(i,i,k)}}{p_{ik}} - \frac{\mathbf{1}_{X=(i,\dagger,k)}}{1-p_{ik}-q_{ik}}\right] & (1 \le j \le 4) \\ \displaystyle\sum_{i=1}^{m-1}\sum_{k=1}^{K} \frac{\partial q_{ik}}{\partial \theta_j}\left[\frac{\mathbf{1}_{X=(i,i+1,k)}}{q_{ik}} - \frac{\mathbf{1}_{X=(i,\dagger,k)}}{1-p_{ik}-q_{ik}}\right] & (5 \le j \le 8) \\ \displaystyle\sum_{k=1}^{K} \frac{\partial f_k^\star}{\partial \theta_j}\left[\frac{\mathbf{1}_{X=(0,1,k)}}{f_k^\star} - \frac{\mathbf{1}_{X=(\bullet,\bullet,k)}}{1-f_k^\star}\right] & (9 \le j \le 10) \end{cases}$$

where, using (6),

$$\frac{\partial f_k^\star}{\partial \theta_j} = \frac{1}{(1+f_k)^2}\frac{\partial f_k}{\partial \theta_j}$$

and the partial derivatives of the vital rates with respect to the parameters θ are given in Appendix A. The non-null elements of the Fisher information matrix then are:

- for $1 \le j, l \le 4$,

$$[I(\theta)]_{jl} = \sum_{i=1}^{m}\sum_{k=1}^{K} \frac{\partial p_{ik}}{\partial \theta_j}\frac{\partial p_{ik}}{\partial \theta_l}\rho_k(1-f_k^\star)d_{ik}\left(\frac{1}{p_{ik}} + \frac{1}{1-p_{ik}-q_{ik}}\right)$$

- for $5 \le j, l \le 8$,

$$[I(\theta)]_{jl} = \sum_{i=1}^{m-1}\sum_{k=1}^{K} \frac{\partial q_{ik}}{\partial \theta_j}\frac{\partial q_{ik}}{\partial \theta_l}\rho_k(1-f_k^\star)d_{ik}\left(\frac{1}{q_{ik}} + \frac{1}{1-p_{ik}-q_{ik}}\right)$$

- for $1 \le j \le 4$ and $5 \le l \le 8$,

$$[I(\theta)]_{jl} = [I(\theta)]_{lj} = \sum_{i=1}^{m}\sum_{k=1}^{K} \frac{\partial p_{ik}}{\partial \theta_j}\frac{\partial q_{ik}}{\partial \theta_l}\rho_k(1-f_k^\star)d_{ik}\frac{1}{1-p_{ik}-q_{ik}}$$

- for $9 \le j, l \le 10$,

$$\begin{aligned} [I(\theta)]_{9,9} &= \frac{1}{\alpha^2}\sum_{k=1}^{K}\rho_k\frac{f_k}{(1+f_k)^2}, \qquad [I(\theta)]_{10,10} = \sum_{k=1}^{K}\rho_k(\mathbf{N}_k.\mathbf{a})^2\frac{f_k}{(1+f_k)^2}, \\ [I(\theta)]_{9,10} &= [I(\theta)]_{10,9} = -\frac{1}{\alpha}\sum_{k=1}^{K}\rho_k(\mathbf{N}_k.\mathbf{a})\frac{f_k}{(1+f_k)^2} \end{aligned}$$

Table 1: Estimates of parameters for the complete model at Paracou: i is the diameter class, $\hat{\mu}_i$, $\hat{\xi}_i$, $\hat{\nu}_i$, $\hat{\kappa}_i$ are the estimates of parameters related to the transition rates of class i, Λ is the likelihood ratio.

i	$\hat{\mu}_i$	$\hat{\xi}_i$		
		Est.	Λ	p-value
1	0.74459	$-2.2896 \cdot 10^{-4}$	194.8	$2.7725 \cdot 10^{-44}$
2	0.75987	$-2.0106 \cdot 10^{-4}$	76.6	$20849 \cdot 10^{-18}$
3	0.72317	$-2.3422 \cdot 10^{-4}$	46.2	$1.0634 \cdot 10^{-18}$
4	0.69263	$-2.7731 \cdot 10^{-4}$	43.2	$5.0081 \cdot 10^{-11}$
5	0.71503	$-2.2683 \cdot 10^{-4}$	21.4	$3.8063 \cdot 10^{-6}$
6	0.73187	$-2.1366 \cdot 10^{-4}$	15.4	$8.6758 \cdot 10^{-5}$
7	0.65946	$-3.2944 \cdot 10^{-4}$	16.6	$4.6497 \cdot 10^{-5}$
8	0.59471	$-4.0799 \cdot 10^{-4}$	17.8	$2.4167 \cdot 10^{-5}$
9	0.50649	$-5.4273 \cdot 10^{-4}$	15.6	$7.7948 \cdot 10^{-5}$
10	0.72583	$-1.9036 \cdot 10^{-4}$	1.1	$3.0394 \cdot 10^{-1}$
11	0.50715	$-6.0784 \cdot 10^{-4}$	25.4	$4.6308 \cdot 10^{-7}$

i	$\hat{\nu}_i$	$\hat{\kappa}_i$		
		Est.	Λ	p-value
1	0.4824	$2.5148 \cdot 10^{-3}$	185.8	$2.6784 \cdot 10^{-42}$
2	0.3369	$2.0196 \cdot 10^{-3}$	62.8	$2.2740 \cdot 10^{-15}$
3	0.3441	$1.8086 \cdot 10^{-3}$	32.5	$1.1676 \cdot 10^{-8}$
4	0.5000	$2.2073 \cdot 10^{-3}$	37.1	$1.1326 \cdot 10^{-9}$
5	0.3471	$0.1542 \cdot 10^{-3}$	17.7	$2.5404 \cdot 10^{-5}$
6	0.3605	$1.6761 \cdot 10^{-3}$	13.6	$2.2503 \cdot 10^{-4}$
7	0.5124	$2.1350 \cdot 10^{-3}$	6.2	$1.2797 \cdot 10^{-2}$
8	0.6541	$2.3094 \cdot 10^{-3}$	10.4	$1.2340 \cdot 10^{-3}$
9	0.3786	$1.4943 \cdot 10^{-3}$	2.8	$9.6238 \cdot 10^{-2}$
10	$3.9772 \cdot 10^{-2}$	$-8.4495 \cdot 10^{-4}$	0.4	$5.0271 \cdot 10^{-1}$

Table 2: Estimates of parameters for the simplified model at Paracou. Λ is the likelihood ratio.

Parameter	Estimate	Λ	p-value
μ_0	0.7445		
μ_1	$-4.7152 \cdot 10^{-3}$		
μ_2	$4.0030 \cdot 10^{-5}$	202.8163	$5.0730 \cdot 10^{-46}$
ξ	$-2.3091 \cdot 10^{-4}$	1354.1400	$1.9411 \cdot 10^{-296}$
ν_0	0.3370		
ν_1	$2.4462 \cdot 10^{-2}$		
ν_2	$-7.9992 \cdot 10^{-4}$	45.9949	$1.1855 \cdot 10^{-11}$
κ	$2.0721 \cdot 10^{-3}$	474.10	$4.1655 \cdot 10^{-105}$

Table 3: Covariance (upper triangle) and correlation (lower triangle) matrix of $\hat{\mathbf{N}}_\infty$.

	$\hat{N}_{\infty 1}$	$\hat{N}_{\infty 2}$	$\hat{N}_{\infty 3}$	$\hat{N}_{\infty 4}$	$\hat{N}_{\infty 5}$	$\hat{N}_{\infty 6}$	$\hat{N}_{\infty 7}$	$\hat{N}_{\infty 8}$	$\hat{N}_{\infty 9}$	$\hat{N}_{\infty 10}$	$\hat{N}_{\infty 11}$
$\hat{N}_{\infty 1}$	51.31	16.27	2.35	−2.67	−3.97	−3.81	−3.17	−2.44	−1.80	−1.29	−0.90
$\hat{N}_{\infty 2}$	0.60	14.28	7.97	2.88	−0.18	−1.60	−2.01	−1.86	−1.48	−1.06	−0.68
$\hat{N}_{\infty 3}$	0.13	0.83	6.53	3.91	1.77	0.39	−0.35	−0.66	−0.72	−0.64	−0.51
$\hat{N}_{\infty 4}$	−0.20	0.42	0.84	3.34	2.37	1.45	0.73	0.24	−0.07	−0.24	−0.30
$\hat{N}_{\infty 5}$	−0.37	−0.03	0.46	0.87	2.22	1.75	1.21	0.72	0.34	0.08	−0.08
$\hat{N}_{\infty 6}$	−0.42	−0.34	0.12	0.62	0.93	1.61	1.27	0.89	0.56	0.29	0.11
$\hat{N}_{\infty 7}$	−0.42	−0.50	−0.13	0.38	0.77	0.95	1.11	0.87	0.63	0.41	0.25
$\hat{N}_{\infty 8}$	−0.39	−0.56	−0.30	0.15	0.56	0.81	0.95	0.76	0.61	0.46	0.33
$\hat{N}_{\infty 9}$	−0.34	−0.53	−0.38	−0.05	0.31	0.60	0.81	0.95	0.54	0.45	0.36
$\hat{N}_{\infty 10}$	−0.28	−0.43	−0.39	−0.20	0.09	0.36	0.61	0.82	0.96	0.41	0.36
$\hat{N}_{\infty 11}$	−0.22	−0.31	−0.35	−0.29	−0.09	0.15	0.41	0.66	0.86	0.97	0.33

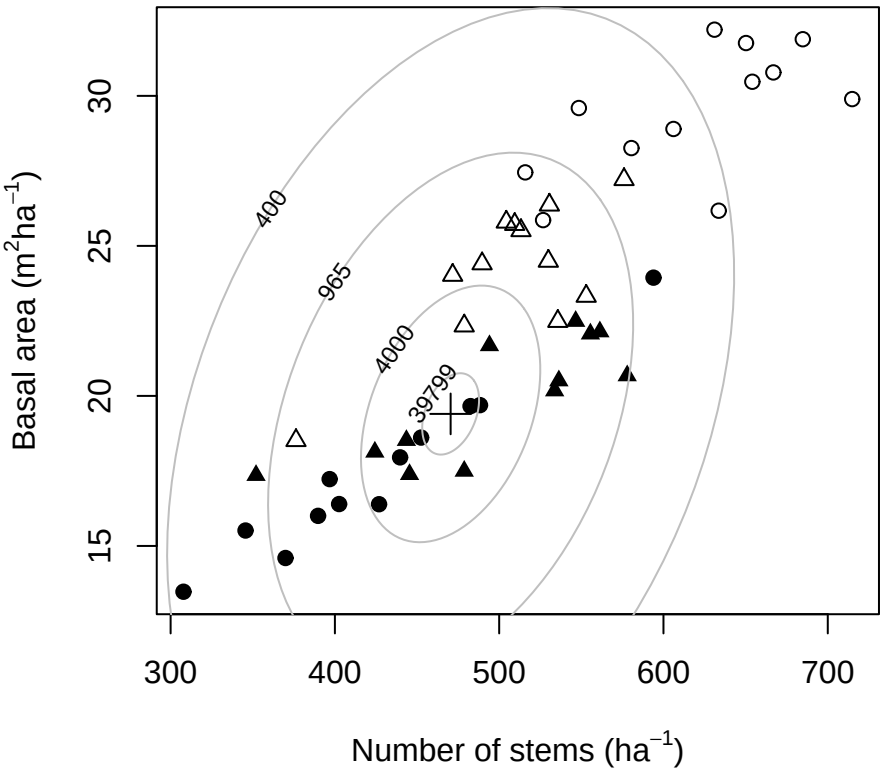


Figure 1: Predicted and observed basal areas and densities at Paracou. Each symbol corresponds to a plot. Black dots are the 12 plots of Paracou that underwent the strongest logging in 1988. Black triangles are the 12 plots of Paracou that underwent moderate logging. White triangles are the 12 plots of Paracou that underwent the lighest logging. White circles are the 12 control plots of Paracou. The cross is the predicted stationary state of a plot according to the simplified model. The ellipses are the confidence ellipses of the predicted state at level 5%, for different sample sizes.

Bibliographie

Alder, D., 1979. A distance-independent tree model for exotic conifer plantations in East Africa. Forest Science 25 (1), 59–71.
URL http://www.bio-met.co.uk/

Alvarez-Buylla, E. R., García-Barrios, R., Lara-Moreno, C., Martínez-Ramos, M., 1996. Demographic and genetic models in conservation biology : applications and perpectives for tropical rain forest tree species. Annual Review of Ecology and Systematics 27, 387–421.

Bergonzini, J. C., Lanly, J. P., 2000. Les forêts tropicales. CIRAD, Montpellier.

Bernardelli, H., 1941. Population waves. Journal of the Burma Research Society 31 (1).

Boos, D., Serfling, R., 1980. A note on differentials and the CLT and LIL for statisticals functions, with application to m-estimates. The Canadian Journal of Statistics 3, 618–624.

Boscolo, M., Kerr, S., Pfaff, A., Sanchez, A., February 1999. What role for tropical forests in climate change mitigation ? The case of Costa Rica. Development Discussion Paper 675, The Harvard Institute for International Development, Harvard University, Cambridge, Massachusetts, USA.
URL http://www.hiid.harvard.edu/pub/ddps/675abs.html

Boscolo, M., Vincent, J. R., 1998. Promoting better logging practices in tropical forests : A simulation analysis of alternative regulations. Development Discussion Paper 652, The Harvard Institute for International Development, Harvard University, Cambridge, Massachusetts, USA.
URL http://www.hiid.harvard.edu/pub/ddps/652abs.html

Boyce, M. S., 1992. Population viability analysis. Annual Review of Ecology and Systematics 23, 481–506.

Buongiorno, J., Gilless, J. K., 2003. Decision Methods for Forest Resource Management. Academic Press.

Caswell, H., 2001. Matrix Population Models : Construction, Analysis and Interpretation, 2nd Edition. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

Chagneau, P., 2006. Optimisation sous contraintes spatiales ; application à la mise en place de parcelles permanentes de suivi des forêts tropicales humides. Mémoire de M2-recherche, Université de Montpellier 2, Montpellier.

Cushing, J. M., 1988. Non-linear matrix models and population dynamics. Natural Resource Modeling 2 (4), 539–580.

Cushing, J. M., Zhou, Y., 1994. The net reproductive number and stability in linear structured population models. Natural Resource Modeling 8 (4), 297–333.

DeAngelis, D. L., Gross, L. J. (Eds.), 1992. Individual-Based Models and Approaches in Ecology - Populations, Communities and Ecosystems. Chapman & Hall, New York.

Doubleday, W. G., 1975. Harvesting in matrix population models. Biometrics 31 (1), 189–200.

Durrieu de Madron, L., Forni, E., Karsenty, A., Loffeier, E., Pierre, J. M., 1998. Le projet d'aménagement pilote intégré de Dimako, Cameroun, 1992-1996. No. 7 in FORAFRI. CIRAD-Forêt, Montpellier, France.

Estève, J., 2001. Étude sur le plan pratique d'aménagement des forêts naturelles de production tropicales africaines : application au cas de l'Afrique Centrale. Premier volet : Production forestière. ATIBT, Paris.

Favrichon, V., 1998. Modeling the dynamics and species composition of tropical mixed-species uneven-aged natural forest : Effects of alternative cutting regimes. Forest Science 44 (1), 113–124.

Favrichon, V., Young Cheol, K., 1998. Modelling the dynamics of a lowland mixed dipterocarp forest stand : Application of a density-dependent matrix model. In : Bertault, J. G., Kadir, K. (Eds.), Silvicultural Research in a Lowland Mixed Dipterocarp Forest of East Kalimantan - The Contribution of STREK Project. CIRAD-Forêt, Montpellier, France, and Ministry of Forestry Research and Development Agency (FORDA), Jakarta, Indonesia, and P.T. Inhutani 1, Jakarta, Indonesia, pp. 229–248.

Fieberg, J., Ellner, S. P., 2001. Stochastic matrix models for conservation and management : a comparative review of methods. Ecology Letters 4 (3), 244–266.

Franc, A., Gourlet-Fleury, S., Picard, N., 2000. Introduction à la modélisation des forêts hétérogènes. ENGREF, Nancy.

Gantmacher, F. R., 1959. Matrix Theory. Chelsea Publishing Company, New York.

Gardiner, C. W., 1985. Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, 2nd Edition. No. 13 in Springer Series in Synergetics. Springer Verlag, Berlin, Germany.

Goodman, L. A., 1967. On the reconciliation of mathematical theories of population growth. Journal of the Royal Statistical Society, Series A 130 (4), 541–553.

Gross, K., Morris, W. F., Wolosin, M. S., Doak, D. F., 2006. Modeling vital rates improves estimation of population projection matrices. Population Ecology 48 (1), 79–89.

Guéneau, S., 2006. Livre blanc sur les forêts tropicales humides. Aanalyses et recommendations des acteurs français. Réponses environnement. La documentation française, Paris.

Hampel, F., 1971. A general qualitative definition of robustness. Journal of American Statistical Association 42, 1887–1896.

Hampel, F., 1986. The Approach Based on Influence Fonctions. Wiley, New York.

Hampel, F. R., 1974. The influence curve and its role in robust estimation. Journal of the American Statistical Association 69, 383–393.

Hauser, C. E., Cooch, E. G., Lebreton, J. D., 2006. Control of structured populations by harvest. Ecological Modelling 196 (3-4), 462–470.

Horn, H. S., 1975. Markovian properties of forest succession. In : Cody, M. L., Diamond, J. M. (Eds.), Ecology and Evolution of Communities. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA, pp. 196–211.

Houde, L., Ledoux, H., 1995. Modélisation en forêt naturelle : stabilité du peuplement. Bois et Forêts des Tropiques 245 (3), 21–26.

Houllier, F., 1986. Échantillonnage et modélisation de la dynamique des peuplements forestiers : application au cas de l'Inventaire Forestier National. Thèse de doctorat, Université Claude Bernard - Lyon I, Lyon, France.

Houllier, F., Lebreton, J. D., Pontier, D., 1989. Sampling properties of the asymptotic behavior of age- or stage-grouped population models. Mathematical Biosciences 95 (2), 161–177.

Huber, P., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics 19, 293–325.

Huber, P. J., 2004. Robust Statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Huston, M. A., DeAngelis, D. L., Post, W. M., November 1988. New computer models unify ecological theory. BioScience 38 (10), 682–691.

Ingram, D., Buongiorno, J., 1996. Income and diversity tradeoffs from management of mixed lowland dipterocarps in Malaysia. Journal of Tropical Forest Science 9 (2), 242–270.
URL `http://www.fpl.fs.fed.us/econ/Publications.htm\#Ingramcd`

Johnson, S. E., Ferguson, I. S., Rong-wei, L., 1991. Evaluation of a stochastic diameter growth model for mountain ash. Forest Science 37 (6), 1671–1681.

Kallianpur, G., Rao, C., 1955. On fisher's lower bound to asymptotic variance of a consistent estimate. Sankhyã 15, 331–342.

Kaye, T. N., Pyke, D. A., 2003. The effect of stochastic technique on estimates of population viability from transition matrix models. Ecology 84 (6), 1464–1476.

Keyfitz, N., 1967. Reconciliation of population models : matrix, integral equation and partial fraction. Journal of the Royal Statistical Society, Series A 130, 61–83.

Lamar, W. R., McGraw, J. B., 2005. Evaluating the use of remotely sensed data in matrix population modeling for eastern hemlock (*Tsuga canadensis* L.). Forest Ecology and Management 212 (1-3), 50–64.

Lande, R., 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. The American Naturalist 142, 911–927.

Lefkovitch, L., 1965. The study of population growth in organisms grouped by stages. Biometrics 21, 1–18.

Legay, J.-M., 1997. L'expérience et le modèle. Un discours sur la méthode. INRA Éditions, Paris.

Leslie, P., 1945. In the use of matrices in certain population mathematics. Biometrika 33, 183–212.

Lewis, E. G., 1942. On the generation and growth of a population. Sankhya 6.

Lotka, 1939. A contribution to the theory of self-renewing aggregates, with special reference to industrial replacement. Annals of Mathematical Statistics 10 (1), 1–25.

Lu, H., Buongiorno, J., 1993. Long- and short-term effects of alternative cutting regimes on economic returns and ecological diversity in mixed-species forests. Forest Ecology and Management 58 (3-4), 173–192.

Menges, E. S., 2000. Population viability analyses in plants : challenges and opportunities. Trends in Ecology and Evolution 15 (2), 51–56.

Michie, B. R., Buongiorno, J., 1984. Estimation of a matrix model of forest growth from re-measured permanent plots. Forest Ecology and Management 8, 127–135.

Morris, W. F., Doak, D. F., 2002. Quantitative Conservation Biology : Theory and Practice of Population Viability Analysis. Sinauer Associates, Inc., Sunderland, MA.

Namaalwa, J., Eid, T., Sankhayan, P., 2005. A multi-species density-dependent matrix growth model for the dry woodlands of Uganda. Forest Ecology and Management 213 (1-3), 312–327.

Nelder, J. A., Mead, R., 1965. A simplex algorithm for function minimization. Computer Journal 7, 308–313.

Neubert, M. G., Caswell, H., 2000. Demography and dispersal : calculation and sensitivity analysis of invasion speed for structured populations. Ecology 81 (6), 1613–1628.

Newnham, R. M., 1964. The development of a stand model for Douglas-fir. Ph.D. Thesis, Faculty of Forestry, University of British Columbia, Vancouver, B. C.

Pavé, A., 1994. Modélisation en biologie et en écologie. Éditions Aléas, Lyon.

Picard, N., Bar-Hen, A., Guédon, Y., 2003. Modelling forest dynamics with a second-order matrix model. Forest Ecology and Management 180 (1-3), 389–400.

Picard, N., Gourlet-Fleury, S., Favrichon, V., 2004. Modelling the forest dynamics at Paracou : the contributions of five models. In : Gourlet-Fleury, S., Guehl, J. M., Laroussinie, O. (Eds.), Ecology and Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana. Elsevier, Paris, pp. 281–296.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 1992. Numerical Recipes in C : The Art of Scientific Computing, 2nd Edition. Cambridge University Press, Cambridge.

Ralston, R., Buongiorno, J., Schulte, B., 2002. WestPro : A Computer Program for Simulating Uneven-aged Douglas-fir Stand Growth and Yield in the Pacific Northwest. Department of Forest Ecology and Management, University of Wisconsin, Madison, WI, USA and USDA Forest Service, Pacific Northwest Experiment Station, Portland, OR, USA.
URL `http://www.forest.wisc.edu/facstaff/buongiorno/book/WestPro/WestPro.htm`

Rieder, H., 1994. Robust Asymptotic Statistics. Springer Verlag.

Rivot, E., 2003. Investigations bayésiennes de la dynamique des populations de saumon atlantique (*Salmo salar* l.) : des observations de terrain à la construction du modèle statistique pour apprendre et gérer. Thèse de doctorat, École Nationale Supérieure Agronomique de Rennes, Rennes, France.

Rogers-Bennett, L., Rogers, D. W., 2006. A semi-empirical growth estimation method for matrix models of endangered species. Ecological Modelling 195 (3-4), 237–246.

Shugart, H. H., Crow, T. R., Hett, J. M., 1973. Forest succession models : a rationale and methodology for modelling forest succession over large regions. Forest Science 19 (3), 203–212.

Sist, P., Picard, N., Gourlet-Fleury, S., 2003. Sustainable cutting cycle and yields in a lowland mixed dipterocarp forest of Borneo. Annals of Forest Science 60 (8), 803–814.

SociétéMathématiqueDeFrance, 1977. Théorie de la Robustesse et Estimation D'un Paramètre : [Séminaire de Statistique Orsay-Paris VII, 1974-1975]. Société Mathématique de France, Paris.

Solomon, A. M., 1986. Transient response of forests to CO_2-induced climate change : simulation modeling experiments in eastern North America. Oecologia (Berlin) 68 (4), 567–579.

Staebler, G. R., 1951. Growth and spacing in an even-aged stand of Douglas fir. Master's Thesis, University of Michigan.

Suzuki, T., Umemura, T., 1974. Forest transition as a stochastic process. In : Fries, J. (Ed.), Growth Models for Tree and Stand Simulation. Proceedings of a S4.01-4 IUFRO Conference. No. 30 in Research Notes. Royal College of Forestry, Department of Forest Yield Research, Stockholm, pp. 358–379.

Taylor, B. L., 1995. The reliability of using population viability analysis for risk classification of species. Conservation Biology 9 (3), 551–558.

Tomé, M., Burkhart, H. E., september 1989. Distance-dependent competition measures for predicting growth of individual trees. Forest Science 35 (3), 816–831.

Usher, M. B., 1966. A matrix approach to the management of renewable resources, with special reference to the selection forests. Journal of Applied Ecology 3, 355–367.

Usher, M. B., 1969. The relation between mean square and block size in the analysis of similar patterns. Journal of Ecology 57, 505–514.

van Hulst, R., 1980. Vegetation dynamics or ecosystem dynamics : dynamic sufficiency in succession theory. Vegetatio 43 (1-2), 147–151.

Vanclay, J. K., 1994. Modelling Forest Growth and Yield - Applications to Mixed Tropical Forests. CAB International, Wallingford.

Vanclay, J. K., 1995. Growth models for tropical forests : A synthesis of models and methods. Forest Science 41 (1), 7–42.

Von Mises, R., 1947. On the asymptotic distributions of differentiable statistical functions. Annals of Statistical Mathematical Statistics 18, 309–348.

Wang, Y., Serfling, R., 2006. Influence functions for a general class of depth-based quantile functions. Journal of Multivariate Analysis 97, 810–826.