

**estimation des volumes
et accroissement
des peuplements forestiers
avec référence particulière aux forêts tropicales
vol. 2 - étude et prévision de la production**

par
d. alder
commonwealth forestry institute, royaume-uni

Les appellations employées dans cette publication et la présentation des données qui y figurent n'impliquent de la part de l'Organisation des Nations Unies pour l'alimentation et l'agriculture aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

M-35
ISBN 92-5-200923-X

Reproduction interdite, en tout ou en partie, par quelque procédé que ce soit, sans l'autorisation écrite de l'Organisation des Nations Unies pour l'alimentation et l'agriculture, seule détentrice des droits. Adresser une demande motivée au Directeur de la Division des publications, Organisation des Nations Unies pour l'alimentation et l'agriculture, Via delle Terme di Caracalla, 00100 Rome, Italie, en indiquant les passages ou illustrations en cause.

© FAO 1980

	<i>Pages</i>
24 Exemples d'expérimentation sur la croissance et la production ..	19
241 Peuplements homogènes	19
241.1 Expériences d'espacement	19
241.2 Expériences d'éclaircie à surface terrière cons- tante	19
241.3 Expériences d'éclaircie avec traitements d'intensité croissante	20
241.4 Expériences factorielles sur les caractéristiques des éclaircies	20
242 Peuplements hétérogènes	21
242.1 Dispositifs en blocs aléatoires	21
242.2 Définition du traitement sylvicole	22
242.3 Caractéristiques des placettes - Mesures à faire ...	22
3 RECUEIL DES DONNEES ET PREMIER TRAITEMENT	23
31 Délimitation des placettes d'échantillonnage	24
311 Localisation	24
312 Identification des placettes permanentes sur le terrain ...	24
313 Limites de la placette et arbres-limites	24
314 Identification des arbres	25
315 Cartographie des arbres de la placette	25
316 Identification des arbres passés à la futaie	26
32 Fiches de mesures pour les placettes d'échantillonnage et premier traitement des données	26
321 Peuplements homogènes	26
322 Forêts hétérogènes	31
323 Première mesure des placettes permanentes	32
33 Analyse des tiges	34
34 Méthodes de mesure de l'accroissement des arbres	35
341 Principe de la mesure	35
342 Mesures successives sur les arbres à contreforts	35
343 Rubans dendrométriques	36
344 Mesures sur les cernes	38
35 Estimation indirecte de la hauteur des arbres et de la hauteur dominante	38
4 SYSTEMES DE STOCKAGE DES DONNEES	40
41 Avantages des systèmes de stockage des données sur ordinateur ..	41
42 Contrôle des données	42
43 Les contrats de fourniture de programmes d'ordinateur	43
44 Systèmes de stockage des données recueillies sur les placettes..	44
441 Introduction	44
442 Structure des fichiers	44
443 Programmes de contrôle et de mise en forme des données	46
444 Résumés de placette	46
445 Autres programmes utilitaires	47
446 Protection de la base de données	47
45 Transfert de données entre systèmes de traitement	47

5	ANALYSE DES DONNEES D'ACCROISSEMENT ET DE PRODUCTION (Peuplements homogènes)	49
51	Classement de la station (détermination de la fertilité)	50
511	Utilisation de la hauteur dominante comme indicateur de fertilité	51
512	Construction des faisceaux des courbes de fertilité	51
512.1	Méthodes graphiques	52
512.2	Méthodes statistiques d'ajustement	56
512.21	La méthode des courbes affines	57
512.22	La méthode du minimum-maximum	59
512.23	Les méthodes de régression sériée	60
512.24	Les méthodes de régression multiple	70
513	Modèles d'évaluation de la fertilité à partir des facteurs du milieu	74
513.1	Modèles fonctionnels de prévision de la classe de fertilité	74
513.2	Choix des variables du milieu à intégrer dans le modèle	75
513.3	Problèmes posés par l'utilisation des modèles d'évaluation de la fertilité	76
52	Méthodes "statiques" de prévision de la production	76
521	Méthodes graphiques de détermination de la relation diamètre moyen/hauteur dominante du peuplement	76
522	Ajustement statistique direct de fonctions donnant le diamètre moyen	79
523	Définition du traitement sylvicole en termes d'âge et de nombre de tiges	81
524	Modèles statiques basés sur la surface terrière ou le volume	82
525	Limites des modèles statiques	83
53	Méthodes "dynamiques" de prévision de l'accroissement et de la production	83
531	La fonction accroissement en surface terrière	84
531.1	L'accroissement en surface terrière en fonction de la hauteur dominante	84
531.2	Autres méthodes de prévision de l'accroissement en surface terrière	86
531.3	Problèmes pratiques posés par l'analyse des données d'accroissement	86
532	Construction d'un modèle d'accroissement sur les résultats d'expériences d'espacement : hypothèse de réponse de MARSH.	86
533	Passage d'un modèle d'accroissement à un modèle de production par intégration	89
533.1	Introduction	89
533.2	Principes de base	89
533.3	Application d'une fonction "production totale" à une sylviculture donnée	91
533.4	Exemple d'utilisation d'une fonction "production totale"	91

	<i>Pages</i>
533.5 Ajustement d'une fonction "accroissement" à des données d'accroissement	94
534 Utilisation d'un modèle d'accroissement en simulation	95
534.1 Eléments nécessaires à la construction d'un modèle de simulation	95
534.2 Méthode de construction d'un modèle de simulation ..	96
534.3 Exemple de modèle de simulation simple	96
54 L'éclaircie	103
541 Le rapport d'éclaircie	103
542 Estimation de la production enlevée en éclaircie (modèles statiques)	104
543 Estimation de la production enlevée en éclaircie (modèles dynamiques)	106
55 La mortalité	106
551 La mortalité à l'installation du peuplement	106
552 Mortalité liée à la densité	107
553 Mortalité liée aux maladies et aux parasites	108
554 Dégâts causés par le vent et le feu	108
56 Prévision du volume	110
561 Tarifs de cubage de peuplement en fonction de la hauteur dominante et de la surface terrière	110
562 Utilisation d'un tarif de cubage d'arbres pour estimer le volume d'un peuplement	111
563 Estimation du volume jusqu'à une découpe déterminée	111
564 Volume des éclaircies	112
57 Techniques avancées de prévision de l'accroissement et de la production	113
571 Modèles par catégories de grosseur	113
572 Modèles de croissance individuelle d'un arbre soumis à la concurrence individuelle de ses voisins	114
6 ANALYSE DES DONNEES D'ACCROISSEMENT ET DE PRODUCTION (PEUPEMENTS HETEROGENES)	115
61 Classement de la station (détermination de la fertilité)	117
62 Modèles de production statiques en forêt hétérogène	120
621 Principes généraux	120
622 Analyse des données	120
623 Choix de la forme de l'équation	121
624 Problèmes posés par les modèles de production statiques dans les peuplements hétérogènes	122
625 Conclusions sur les modèles de production statiques dans les peuplements hétérogènes	123
63 Modèles utilisant des matrices de transition	123
631 Définition	123
632 Méthodes de construction des matrices de transition	126
632.1 A partir de données où les arbres sont suivis individuellement	126
632.2 A partir de données décrivant le peuplement par catégories de grosseur	127

	<i>Pages</i>
633 Modèles de transition dans les peuplements hétérogènes	129
634 Inconvénients des modèles de transition	130
64 Modèles de croissance individuelle d'un arbre soumis à la concurrence du peuplement	131
641 Définition	131
642 Variables dynamiques et variables allométriques	131
643 Modélisation de la concurrence	131
644 Données nécessaires et analyse de données	132
645 Structure du modèle	133
646 Passage à la futaie, mortalité et exploitation	134
647 Conclusions sur les modèles de croissance individuelle	135
7 CONTROLE DE LA VALIDITE DES MODELES D'ACCROISSEMENT ET DE PRODU- TION	136
71 Nécessité de ce contrôle	137
72 Les données-tests	138
73 Les écarts résiduels	139
74 Comparaisons graphiques	140
75 Définition des limites d'utilisation du modèle	140
8 UTILISATION DES MODELES POUR L'AMENAGEMENT	141
81 Introduction	142
82 Peuplements équiennes	142
83 Peuplements hétérogènes	143
A N N E X E S	
.....	
<u>ANNEXE A</u> : TECHNIQUES STATISTIQUES ET MATHÉMATIQUES UTILISÉES DANS LES ÉTUDES DE PRODUCTION FORESTIÈRE	
	145
A1 COURBES ET ÉQUATIONS	146
A11 La droite	146
A12 Transformation des variables pour l'ajustement d'une équation à une courbe	147
A13 Exemple d'ajustement avec transformation de la courbe en droite.	148
A14 Approximation d'une courbe par un polynôme	148
A15 Relations entre 3 variables	153
A16 Exemple : ajustement d'une équation à un système de courbes par harmonisation	154
A2 AJUSTEMENT D'UNE ÉQUATION PAR LA MÉTHODE DES MOINDRES CARRÉS	158
A21 Régression linéaire simple	158
A22 Régression linéaire à deux variables explicatives	162
A23 Transformation des données et ajustement d'une équation	165
A24 La régression linéaire multiple	170
A25 Estimation des paramètres dans l'ajustement du modèle de régres- sion linéaire multiple	171

	<i>Pages</i>
A26 Analyse des résidus	182
A27 Régression linéaire pondérée	185
A28 Comparaison d'équations obtenues par régression	187
A29 Régression sériée avec emploi de variables conditionnelles	194
A3 RESOLUTION DES EQUATIONS	199
A31 Résolution de l'équation du 2ème degré	199
A32 Résolution graphique des équations	199
A33 Résolution numérique des équations	201
A4 AJUSTEMENT DES MODELES NON LINEAIRES	204
<u>ANNEXE B</u> : TABLES DES FONCTIONS COURANTES UTILISEES DANS LES MODELES DECRITS DANS LE TEXTE	210
<u>ANNEXE C</u> : FICHES DE CALCUL ET DE RECUEIL DES DONNEES DECRITES DANS LE TEXTE	215
<u>ANNEXE D</u> : BIBLIOGRAPHIE COMMENTEE	224

PREAMBULE

Il est évident pour les régisseurs forestiers que la possibilité d'estimer le volume des arbres et des peuplements et de prédire ce que la forêt produira selon les stations en fonction des différents types de traitement sylvicole est indispensable à tout processus rationnel de planification forestière. Il y a cependant une diversité considérable d'opinions à propos du concept de "production", des façons de l'estimer et d'en prédire l'évolution.

Ce manuel cherche à codifier les pratiques courantes d'estimation et de prédiction du volume des arbres et des peuplements, afin qu'elles puissent être mises en œuvre par les personnes chargées de ce type de travail sans qu'elles soient pour autant spécialistes dans ce domaine.

Il est cependant nécessaire de souligner que, s'agissant d'un domaine en pleine évolution, notamment dans les tropiques, ce manuel doit être considéré comme provisoire et sujet à de futurs ajustements concernant soit des situations nouvelles et particulières, soit des techniques nouvelles, soit des techniques existantes mais non mentionnées dans le texte et qui pourraient se révéler supérieures.

C'est pourquoi il ne s'agit pas d'un manuel au sens vrai du terme mais davantage d'un recueil de conseils concernant le choix des méthodes auxquelles on a joint des instructions précises dans certains cas.

Ce manuel se réfère plus particulièrement aux forêts tropicales, tant naturelles que artificielles. Compte tenu de la difficulté d'évaluer la croissance et la production des peuplements naturels hétérogènes et non équiennes, les méthodes indiquées pour la construction des modèles de croissance s'appliquent essentiellement aux forêts équiennes monospécifiques. En ce qui concerne les forêts hétérogènes aucune instruction précise n'est donnée mais un certain nombre d'approches possibles sont proposées.

Le manuel comprend deux tomes. Le premier tome décrit les techniques de mesure et de calcul du volume des arbres et des peuplements. Le second concerne la croissance et la production. Une description des techniques statistiques et mathématiques, des tables statistiques sélectionnées, des exemples d'imprimés de calcul et de saisie de données et une bibliographie commentée sont inclus dans une série d'annexes.

Le premier tome du manuel a été rédigé par Francis Cailliez du Centre Technique Forestier Tropical (CTFT) à Nogent-sur-Marne (France) et le second par Danis Alder du Commonwealth Forestry Institute (CFI) à Oxford (Grande-Bretagne) qui a également réuni les annexes. Le travail des deux auteurs a été coordonné par Joran Fries, Université des sciences agronomiques, Uppsala (Suède). Le travail a été conçu et guidé par Jean-Paul Lanly et Karn Deo Singh de la Division des ressources forestières de la FAO. Jean Clément (CTFT) a été associé à la phase initiale de l'étude.

La traduction française du deuxième tome et des annexes a été réalisée par Martine et Pierre Duplat. Excepté les références d'articles techniques cités par l'auteur, toutes les notes en bas de page sont rédigées par les traducteurs. La frappe a été effectuée par Catherine Gueguen, du CTFT.

La première version du manuel fut présentée à la réunion du groupe de recherche S4.01 de l'IUFRO (dendrométrie, croissance et production) qui s'est tenue à Oxford en septembre 1979 et elle fut discutée en détail pendant une journée complète. La plupart des participants étaient des spécialistes de dendrométrie des pays tropicaux invités spécialement par la FAO pour faire une analyse critique du manuel. Il avait été également envoyé à certains autres experts aux fins de commentaire. Les auteurs ont ensuite revu le texte pour tenir compte des remarques formulées.

Ce manuel, qui est le premier du genre dans le domaine de la foresterie tropicale, laisse le champ libre pour des améliorations et compléments futurs, particulièrement dans le cas des peuplements hétérogènes non équiennes pour lesquels des études complémentaires sont nécessaires. Toutes les suggestions à ce sujet seront particulièrement appréciées.

M.A. Flores Rodas
Sous-Directeur général
Département des forêts

CHAPITRE 1

INTRODUCTION :
LA PREVISION DE L'ACCROISSEMENT
ET DE LA PRODUCTION

1 INTRODUCTION : LA PREVISION DE L'ACCROISSEMENT ET DE LA PRODUCTION

11 POURQUOI PREVOIR L'ACCROISSEMENT ET LA PRODUCTION ?

Si l'on veut comprendre pourquoi les méthodes de prévision de l'accroissement ⁽¹⁾ et de la production ⁽²⁾ des peuplements forestiers sont si diverses, il est bon d'examiner de plus près à quoi peut servir cette prévision.

111 Aménagement forestier et production

Une gestion forestière est efficace quand le traitement appliqué assure une évolution du peuplement telle que sa valeur économique et/ou sociale augmente plus vite que le coût actualisé du traitement.

Toute coupe a une incidence, plus ou moins grande, sur l'avenir du peuplement sur pied : un taux de récolte trop élevé peut se solder par l'épuisement total de la ressource forestière, tandis qu'un prélèvement trop faible peut à la fois priver une communauté de ressources immédiates et réduire la capacité d'accroissement de la forêt.

Il est évident que l'on ne peut décider rationnellement du traitement, de l'intensité et de la rotation des coupes, que si l'on est capable de quantifier la réponse de la forêt à ces opérations. C'est à quoi visent les études d'accroissement et de production.

112 Recherche sylvicole et aménagement

L'objectif principal des études d'accroissement et de production est donc de quantifier la réponse de la production forestière aux traitements subis, mais il ne faut pas oublier pour autant qu'il existe une forte interaction entre ces études et des formes plus qualitatives de recherche sylvicole :

- (i) La connaissance que l'on a du tempérament d'une essence aide à choisir le type de modèle à utiliser pour représenter sa croissance ; dans les forêts complexes, cette connaissance permet de regrouper certaines essences sur des bases logiques. Les caractéristiques sylvicoles d'une essence suggéreront les caractéristiques du modèle correspondant, c'est-à-dire le choix des relations qui le constituent.
- (ii) Inversement, on peut utiliser un modèle quantitatif, s'il est correct, pour tester des hypothèses sylvicoles et concevoir les traitements et dispositifs expérimentaux susceptibles de fournir des résultats intéressants.

⁽¹⁾ en anglais = growth

⁽²⁾ en anglais = yield

113 Recherche écologique et gestion de l'environnement

Des modèles quantitatifs conçus pour prévoir l'accroissement et la production de peuplements forestiers peuvent contribuer à la recherche écologique et à la gestion de l'environnement de plusieurs façons :

- (i) Un modèle forestier peut servir à déterminer la quantité de lumière parvenant au sol, à différents stades du cycle de croissance.
- (ii) On peut assez facilement rendre un modèle forestier capable d'indiquer la biomasse et le taux de production du peuplement forestier.
- (iii) La connaissance de la période du cycle de croissance pendant laquelle le peuplement est accessible aux grands herbivores est très utile pour la gestion de la faune sauvage.

Les techniques utilisées en modélisation écologique sont en général assez différentes de celles utilisées en foresterie pour les études d'accroissement et de production. Ces dernières doivent en effet s'attacher à prévoir de façon très précise les caractéristiques géométriques du peuplement, alors qu'en écologie on peut traiter comme un tout chaque population et chaque niveau d'un écosystème.

En outre, les modèles écologiques se consacrent généralement à la description ou à l'explication des traits qualitativement importants d'un écosystème, sans essayer d'atteindre un niveau élevé de précision, d'ailleurs rarement possible ou nécessaire. En revanche, les modèles forestiers doivent être relativement précis s'ils veulent atteindre leur but et donc justifier leur existence.

Cependant, il y a des points de contact entre les techniques de la modélisation écologique et celles de la modélisation forestière :

- (i) Dans les forêts hétérogènes, la composition en essences peut être modélisée par des techniques de dynamique des populations.
- (ii) On peut quantifier la mortalité, les altérations du bois ou les pertes d'accroissement dues à une maladie ou une attaque parasitaire d'étiologie connue, à l'aide d'un modèle de dynamique de population dans lequel :
 - . l'information sur "l'habitat" est tirée d'un modèle forestier,
 - . l'accroissement des arbres est influencé par le niveau de population du parasite.

Dans l'avenir, ces points de contact se multiplieront probablement. On pourra notamment aborder avec profit, grâce à des formes modifiées de modèles écologiques du type flux d'énergie/cycle des matières nutritives, les difficiles problèmes que pose la forêt hétérogène, d'essences et d'âges variés, comme ceux que soulève l'intérêt croissant porté aux produits de la forêt autres que le bois.

12 METHODOLOGIE DE LA PREVISION DE L'ACCROISSEMENT ET DE LA PRODUCTION

On peut y distinguer quatre étapes, chacune faisant l'objet d'un des paragraphes suivants.

121 Mesure de l'accroissement et de la production

Lors des observations qui serviront de base à la construction du modèle, la mesure de l'accroissement et de la production soulève deux types de problèmes. Le premier est la définition de ce qui constitue la production : s'agira-t-il du volume de bois d'oeuvre récolté, du volume de bois de certaines essences seulement, ou de divers produits autres que le bois comme l'écorce, les feuilles, la résine, ... etc ? Dans les pays tropicaux, on s'intéresse généralement au volume et à la composition de l'ensemble du bois produit : bois d'oeuvre, bois de trituration, perches, bois de feu. Dans les peuplements hétérogènes, on ne peut pas parler de la production en ignorant la composition en essences, car celle-ci a une influence sur les emplois possibles du produit.

Le deuxième problème est posé par la réalisation des mesures. Une fois que l'on a défini clairement ce qui doit être mesuré, les principales difficultés sont d'ordre pratique : accès à la forêt, délimitation et mesure des placettes, maintenance de placettes permanentes sur de longues périodes. Ces points seront examinés dans les chapitres 2 et 3.

122 Construction d'un modèle mathématique, et ajustement aux données d'accroissement et de production

Un modèle mathématique est un jeu d'équations (ou de courbes) exprimant les relations qui existent entre les variables qu'on y a introduites. Il doit être "ajusté" à des données tirées de l'observation de la réalité.

Le procédé d'ajustement du modèle peut être statistique et utiliser par exemple la régression linéaire, ou subjectif et comporter le tracé à la main de courbes dans les nuages de points traduisant les données. Il est d'ailleurs possible de passer, si besoin est, d'une courbe tracée à la main à une équation : l'annexe A décrit plusieurs méthodes d'ajustement de courbe qui le permettent.

Le choix du type de courbe tracée ou d'équation ajustée peut s'appuyer sur une loi de croissance "naturelle", ou être purement empirique. Dans ce cas, on choisit un type de courbe ou d'équation uniquement pour son aptitude à rendre compte d'une forme donnée.

Actuellement, il n'existe pas vraiment de fonction représentant l'accroissement des arbres qui soit basée sur une loi naturelle bien que certaines, comme la fonction de CHAPMAN-RICHARDS, puissent être considérées comme reflétant partiellement des processus de croissance.

123 Contrôle de la validité du modèle

Une fois le modèle construit et ajusté aux données, on doit le tester pour contrôler sa validité et sa précision. Le mieux est de le faire sur un second jeu de données, n'ayant pas servi à l'ajustement des fonctions du modèle : on fait prévoir au modèle le comportement des peuplements qui ont fourni les données-tests et on compare les résultats aux observations réelles.

Il est souvent nécessaire de répéter plusieurs fois ce processus de validation, en apportant chaque fois au modèle les adaptations ou les corrections qui s'imposent pour résoudre les anomalies mises en évidence.

On peut citer plusieurs raisons au mauvais fonctionnement d'un modèle lors du contrôle. Il se peut :

- (i) que le jeu original de données et le jeu test soient représentatifs de schémas de croissance différents,
- (ii) qu'on ait employé, dans la construction du modèle, des méthodes d'ajustement inadéquates,
- (iii) qu'on ait extrapolé, pour le contrôle sur le second jeu de données, certaines fonctions dans un domaine où elles sont imprécises,
- (iv) que le modèle, mettant en jeu un système d'équations, devienne instable quand on l'utilise comme un tout, même si chacune des fonctions s'ajuste bien aux données,
- (v) qu'une erreur humaine quelconque se soit produite pendant la transcription ou l'utilisation des différentes courbes ou équations du modèle.

On a développé un peu longuement ces quelques points pour insister sur la nécessité de tester soigneusement tout modèle avant de l'appliquer à l'aménagement ou à la recherche. Le chapitre 7 traite en détail de la validation des modèles.

124 Utilisation du modèle

On peut utiliser un modèle d'accroissement et de production de trois façons différentes :

- (i) sous forme d'une simple table (ou courbe) ou d'un jeu de tables (ou courbes). Celles-ci peuvent alors être utilisées directement par les aménagistes, ou entrées dans un ordinateur pour la mise à jour des données d'un inventaire,
- (ii) sous forme d'un programme pour ordinateur ou calculateur, qui peut alors éditer une table (ou des courbes) de production pour une sylviculture donnée. Cette formule présente un grand intérêt quand le modèle est assez flexible pour qu'il soit impossible de rendre compte dans un seul jeu de tables de toutes les prévisions qu'il permet de faire,
- (iii) sous forme d'un sous-programme constituant un sous-modèle dans un programme d'ordinateur pour l'aménagement forestier ; en général, celui-ci comprendra aussi une base de données d'inventaire et des contraintes techniques ou économiques sur la sylviculture et l'exploitation.

CHAPITRE 2

DISPOSITIFS D'ETUDE
DE LA PRODUCTION

2 DISPOSITIFS D'ETUDE DE LA PRODUCTION

21 CONSEQUENCES DU CHOIX D'UNE METHODE D'AJUSTEMENT PAR REGRESSION

Le choix des dispositifs d'échantillonnage ou d'expériences pour les études d'accroissement et de production dépend du type de modèle que l'on veut ajuster aux données qu'ils fourniront. Il s'agit en général d'un modèle de régression ; il faut alors garder présents à l'esprit les points suivants :

- (i) Si l'on sait que le modèle à ajuster est de forme linéaire, on intensifiera l'échantillonnage ou les expériences aux deux extrémités (ou aux bords du domaine de variation dans le cas de plusieurs variables explicatives).
- (ii) Le plus souvent, on ne connaît pas la forme précise de la fonction à ajuster, vraisemblablement non linéaire. On a alors intérêt à diviser le domaine de variation de la variable explicative en 5 classes égales et à échantillonner avec la même intensité dans chacune.
- (iii) Dans chaque classe du domaine de variation, il est souhaitable que l'intensité d'échantillonnage soit proportionnelle à la variance des échantillons autour du modèle ; ceci s'applique tout particulièrement à la prévision du volume (*cf. première partie de ce manuel*).
- (iv) L'échantillonnage au hasard ou systématique par zone, qui convient particulièrement à l'inventaire forestier, ne fournit pas une base efficace à la construction des modèles d'accroissement et de production. Il entraîne en effet une intensité d'échantillonnage trop grande dans la partie centrale et trop faible sur les bords du domaine de variation.
- (v) Les plans d'expérience pour les études d'accroissement et de production doivent toujours inclure des modalités extrêmes, en particulier pour la densité du peuplement : le modèle que l'on ajustera aux données ainsi obtenues en sera d'autant plus précis.

22 ECHANTILLONNAGE EN VUE DE LA CONSTRUCTION D'UN MODELE

L'échantillonnage de la réalité constitue une solution de remplacement à l'expérimentation quand il est impossible au chercheur de contrôler toutes les variables entrant dans le modèle. Dans les études de production, c'est surtout le cas de la variable "station", puisqu'on peut contrôler le type de peuplement en implantant l'expérience dans un peuplement du type désiré ou même en créant un peuplement de ce type, puisqu'on peut contrôler la densité du peuplement par des opérations sylvicoles, ... etc.

Il est généralement plus efficace, c'est-à-dire moins coûteux pour une précision donnée du modèle, d'installer des dispositifs expérimentaux que de recueillir les données par échantillonnage ; mais on a besoin des deux types de données si on veut que le modèle prenne effectivement en compte les variations de station.

221 Placettes temporaires

On utilise en principe les placettes temporaires ⁽¹⁾ pour la détermination de relations où le temps n'intervient pas. En fait, quand les cernes annuels sont discernables, leur observation permet l'étude de relations qui font intervenir le temps.

221.1 Inventaire forestier

Les dispositifs d'inventaire forestier sont conçus avant tout pour fournir une estimation précise du peuplement sur pied sur une certaine surface, mais une partie de l'information recueillie peut être utilisée dans une étude de production. On trouvera dans le Manuel d'Inventaire Forestier de la FAO l'exposé général des problèmes posés par la conception, la réalisation et l'analyse des résultats d'un inventaire forestier.

Ce serait en général une perte de temps que de mesurer sur toutes les placettes d'un inventaire les variables intéressantes pour l'étude de production ; il vaut mieux consacrer ce temps à des mesures plus détaillées sur un sous-échantillon.

221.2 Estimation de l'accroissement à partir des cernes annuels

Quand les cernes annuels sont bien visibles, on peut remplacer le suivi de placettes permanentes par l'étude de placettes temporaires. L'estimation de l'accroissement à partir des cernes annuels est généralement plus difficile et plus coûteuse qu'à partir de placettes permanentes, mais elle est évidemment plus rapide.

221.3 Echantillonnage pour la détermination de relations allométriques

Une relation allométrique est une relation entre deux grandeurs observées sur un même arbre : par exemple, diamètre du houppier et diamètre du fût, ou hauteur totale et hauteur du fût, ... ; ce genre de relations peut jouer un rôle important dans certains modèles. Comme les inventaires forestiers ne fournissent généralement pas les données nécessaires sous une forme adéquate, il faut réaliser un échantillonnage particulier pour déterminer la relation cherchée.

L'unité d'échantillonnage ⁽²⁾ est normalement l'arbre, mais il peut être plus commode d'installer des placettes et d'en mesurer tous les arbres. Le nombre d'échantillons à mesurer dépend de la relation à établir ; on peut procéder de la façon suivante : analyser les données au fur et à mesure de la réalisation de l'échantillon et cesser les mesures quand on a atteint le degré de précision désiré.

Un tarif de cubage d'arbre ⁽³⁾ constitue un exemple particulier de relation allométrique.

(1) en anglais = temporary plots

(2) en anglais = sampling unit

(3) en anglais = tree volume tariff

221.4 Echantillonnage pour l'étude des caractéristiques des coupes

Dans la plupart des modèles de prévision de la production, les coupes d'éclaircie ou de jardinage sont décrites par des spécifications théoriques. On peut toujours supposer que la coupe sera réalisée exactement comme prévu, mais on peut aussi examiner, sur un échantillon, la relation qui existe entre les spécifications théoriques et les résultats obtenus.

En outre, la plupart des modèles ont besoin d'informations sur la coupe qui ne sont pas explicitement contenues dans les spécifications ; il peut être nécessaire, par exemple, de connaître la distribution en diamètre des tiges enlevées, ou la relation entre nombre de tiges et surface terrière enlevées.

Un échantillonnage réalisé peu après l'exploitation peut fournir de l'information sur ces points, mais on peut aussi planter de façon invisible des placettes semi-permanentes qui seront mesurées avant et après l'exploitation pour obtenir un inventaire précis des arbres enlevés.

Les études sur ce sujet requièrent des placettes plus grandes que les autres études liées à la production car, dans la pratique, une exploitation a des effets localement très hétérogènes (chemins de tirage, aires de chargement,...). On peut conseiller les ordres de grandeur suivants :

- dans un peuplement homogène : 0,1 à 0,5 ha
- en forêt tropicale hétérogène : 5 à 10 ha

221.5 Etude de la régénération

Dans les forêts tropicales hétérogènes, dans les futaies irrégulières, ou dans les futaies régulières régénérées naturellement, il peut être très intéressant que le modèle de prévision de la production comporte aussi des renseignements sur la régénération.

Les placettes d'étude de la régénération sont nécessairement petites. Il peut s'agir soit de sous-placettes incluses dans des placettes d'inventaire forestier classique, soit de placettes d'un dispositif particulier d'échantillonnage réalisé 3 à 5 ans après l'exploitation. Le plus souvent, on subdivise ces placettes en petits carrés ⁽¹⁾, et on note sur chacun d'eux la présence ou l'absence de semis de chaque essence ; on peut aussi compter les semis dépassant une certaine taille, mais ceci ajoute peu à l'utilité de l'information. Les surfaces généralement adoptées sont 0,01 ha (10 x 10 m) ou 0,04 ha (20 x 20 m) pour les placettes, et 1 m² ou 4 m² pour les petits carrés.

222 Placettes permanentes

La plupart des forestiers considèrent qu'un échantillonnage de placettes permanentes ⁽²⁾ est la principale source de données pour la construction d'un modèle d'accroissement et de production. C'est souvent vrai, mais il ne faut pas oublier que des expérimentations peuvent s'avérer nécessaires, ne serait-ce que pour introduire dans le domaine d'étude des modalités extrêmes que l'on ne trouve pas dans la forêt ; d'un autre côté, la possibilité de compter les cerne annuels peut conduire à se contenter de placettes temporaires.

⁽¹⁾ en anglais = quadrats

⁽²⁾ en anglais = permanent sample plots

222.1 Nombre de placettes permanentes nécessaire

On ne peut pas définir le nombre de placettes permanentes nécessaire à partir de critères purement statistiques. La précision réelle du modèle ajusté aux données qu'elles fournissent dépend autant de la localisation des placettes et du temps pendant lequel on les a suivies que des covariances entre variables explicatives ou entre coefficients du modèle ajusté.

L'expérience suggère cependant qu'une centaine de placettes bien réparties dans l'ensemble des types de station et de traitement sylvicole peut suffire pour un type de forêt donné ou pour une essence en plantation, sauf s'il y a visiblement un schéma de croissance différent dans une partie de la région étudiée.

222.2 Choix de l'emplacement des placettes permanentes

Les placettes permanentes doivent être installées en nombre égal dans les stations :

- pauvres,
- moyennement fertiles,
- riches,

dans les peuplements :

- clairs,
- de densité moyenne,
- denses,

et dans les peuplements :

- jeunes (ou récemment exploités, pour les forêts hétérogènes),
- d'âge moyen (ou entre deux exploitations,...),
- proches de l'âge d'exploitabilité (ou avant exploitation,...).

On aura ainsi une distribution des placettes qui sera très irrégulière dans l'espace et qui sous-représentera les peuplements moyens ; mais, comme nous l'avons vu au paragraphe 21, c'est la méthode d'échantillonnage la plus efficace pour estimer les paramètres d'une régression.

Une stratification de ce type sera parfois irréalisable, par manque de connaissance sur les conditions de croissance dans la région ; on peut alors implanter les placettes systématiquement, ou en utilisant une stratification géographique, pour assurer une couverture uniforme de la région, mais il en faudra un nombre beaucoup plus important qu'avec la stratification efficace décrite ci-dessus.

222.3 Taille des placettes permanentes

En règle générale, c'est le type de forêt, l'hétérogénéité du peuplement et de la distribution des essences, qui commandent le choix de la taille des placettes permanentes d'échantillonnage.

Dans les forêts tropicales hétérogènes, une surface de 1 ha convient généralement ; il est souvent commode de la diviser en carrés de 10 × 10 m. Dans les peuplements homogènes, on se contente souvent d'une surface de l'ordre de 0,05 ha. On peut évidemment modifier ces chiffres pour tenir compte des circonstances, et notamment les augmenter dans les peuplements âgés.

Pour les placettes d'expérience, on adoptera normalement des surfaces plus grandes. Le tableau ci-dessous donne quelques détails :

Taille des placettes permanentes

Type de forêt	hétérogène	homogène
Placettes d'échantillonnage	1 à 2 ha	0,04 à 0,08 ha
Placettes d'expérience (sans les bordures)	1 à 5 ha	0,08 à 0,12 ha
Etude d'exploitations réelles ⁽¹⁾	5 à 10 ha	0,10 à 0,50 ha

mais ces chiffres ne sont pas à prendre comme des règles strictes.

222.4 Forme des placettes permanentes

D'une façon générale, on donne aux placettes permanentes une forme rectangulaire ou circulaire. Dans les inventaires, on utilise parfois d'autres formes telles que croix ou grappe de placettes circulaires qui présentent des avantages spécifiques pour l'échantillonnage d'une surface, mais n'ont pas d'intérêt particulier pour des placettes permanentes.

Les placettes circulaires sont plus rapides à installer que les placettes rectangulaires tant que leur surface ne dépasse pas 0,10 ha dans les peuplements clairs, 0,05 ha dans les peuplements denses. En outre, leur emploi dans les plantations permet d'éviter les problèmes de représentativité de la surface dus à la disposition des plants en lignes.

En revanche, quand la surface des placettes doit dépasser 0,10 ha, il vaut mieux adopter une forme rectangulaire. On choisira le rapport longueur/largeur en fonction des circonstances ; sur les pentes fortes, on a intérêt à prendre un rapport élevé (jusqu'à 5), avec la longueur dans le sens de la pente ; en terrain horizontal, une placette carrée a un périmètre plus petit et sera donc plus facile à délimiter et à mesurer.

222.5 Périodicité et date des mesures

L'intervalle normal entre deux mesures d'une placette permanente dépend de la vitesse de croissance des arbres. Il peut être intéressant de remesurer après un intervalle plus court une placette que l'on vient d'installer, pour profiter le plus rapidement possible des données d'accroissement de cette nouvelle placette.

Il faut noter que, d'une façon générale, la précision sur les accroissements des arbres est d'autant meilleure que l'intervalle entre les mesures est plus long.

(¹) par opposition aux exploitations simulées qui entrent dans la catégorie des expériences.

Un organisme responsable d'un grand nombre de placettes permanentes aura intérêt à planifier les mesures successives de façon à ne jamais avoir à supporter, une année donnée, la mesure de plus du tiers des placettes.

On peut fixer approximativement l'intervalle entre deux mesures de la façon suivante ⁽¹⁾ :

Type de forêt	Périodicité des mesures
Jeune plantation en zone tropicale	1 an
Plantation plus âgée ou autres forêts homogènes en zone tropicale	2 à 4 ans
Forêt tropicale hétérogène	3 à 5 ans
Forêt tempérée homogène ...	3 à 5 ans

On doit évidemment fixer la date des mesures successives en tenant compte des effets saisonniers. S'il y a une saison de végétation bien définie, les mesures seront faites après la fin de cette saison, à une date quelconque pendant le repos de la végétation. Si l'on mesure une placette à intervalle d'un an, la mesure doit être faite toujours le même mois pour fournir une estimation de l'accroissement annuel et permettre les comparaisons. Avec des intervalles plus longs et des climats moins saisonniers, la fixation précise de la date de mesure devient moins importante.

Dans certains cas, le choix de la date peut être conditionné par les possibilités d'accès à la forêt ou d'emploi d'une main d'oeuvre saisonnière.

222.6 Echantillonnage avec renouvellement partiel

Au sens strict, l'échantillonnage avec renouvellement partiel ⁽²⁾ est un dispositif d'inventaire continu, dans lequel des placettes semi-permanentes apportent un supplément à l'information tirée de placettes temporaires. On peut cependant appliquer ce concept aux placettes permanentes d'échantillonnage.

Au fil des mesures successives, une placette permanente apporte de moins en moins d'information par rapport à la mesure précédente, à moins qu'elle ne soit installée dans une strate âge x station x densité du peuplement qui n'a pas été bien échantillonnée.

(¹) sans oublier qu'il est bon de faire une mesure immédiatement avant et une autre immédiatement après toute opération sylvicole se traduisant par une coupe.

(²) en anglais = sampling with partial replacement.

Dans les peuplements équiennes, on peut adopter deux stratégies différentes d'échantillonnage en placettes permanentes :

- (i) On implante des placettes dans toutes les classes d'âge ; l'échantillonnage est alors plus efficace si on renouvelle une partie des placettes après la quatrième ou cinquième mesure.
- (ii) On n'implante des placettes que dans des jeunes plantations (parce qu'il n'y a pas encore d'autres classes d'âge). Il faut alors suivre une partie des placettes, disons un tiers, tout au long de la révolution ⁽¹⁾, les deux autres tiers étant renouvelées après trois ou quatre mesures.

Dans les forêts hétérogènes, on se trouve devant une situation analogue, mais en remplaçant l'âge par le nombre d'années écoulées depuis la dernière exploitation.

23 DISPOSITIFS EXPERIMENTAUX

L'expérimentation constitue la source de données la plus efficace et la plus utile pour la construction de modèles d'accroissement et de production. Cependant, l'utilité de nombreuses expériences entreprises en recherche forestière est limitée par une incapacité à concevoir clairement le modèle mathématique que l'expérience doit servir à tester ou à paramétrer.

Les études d'accroissement et de production n'ont pas pour but principal de mettre en évidence des différences significatives entre divers traitements, mais de construire des courbes de réponse. Il ne faut pas croire non plus que des études forestières à long terme puissent apporter une solution à des problèmes qui trouvent leur origine dans les conditions économiques du moment : on s'apercevrait alors, au moment où l'expérience commence à fournir des données utiles, que les résultats ne sont plus pertinents parce que les conditions économiques ont changé.

Il faut donc entreprendre des expériences en allant à la recherche de principes généraux, exprimés sous la forme d'un modèle mathématique, plutôt que pour sélectionner "la meilleure" des modalités expérimentées. En outre, les variables à mesurer ne seront pas définies en termes économiques, mais toujours en termes écologiques ou sylvicoles. Bien sûr, on aura toujours besoin d'expérimenter à court terme sur des variables économiques pour chiffrer le coût des traitements sylvicoles, déterminer la production utilisable en fonction des méthodes d'exploitation..., mais ces questions sortent du cadre de ce manuel.

On peut diviser les principaux dispositifs expérimentaux utilisables dans la recherche sur la croissance et la production en deux groupes : les dispositifs aléatoires et les dispositifs systématiques. Les premiers peuvent être soumis à une analyse de variance traditionnelle tandis que les seconds, généralement moins coûteux à réaliser, conviennent quand on a prévu de traiter les données par régression ; mais, les données tirées de dispositifs aléatoires pouvant aussi être traitées par régression, ceux-ci sont probablement préférables dans tous les cas, sauf dans les expériences d'espacement en plantation.

(1) en anglais (attention aux confusions) : rotation.

231 Dispositifs aléatoires

La caractéristique essentielle des dispositifs aléatoires ⁽¹⁾ est la suivante : chaque placette de l'expérience se voit attribuer un "traitement" ⁽²⁾ (c'est-à-dire une des modalités expérimentées) par un procédé de choix au hasard ; on utilise en général une table de nombres au hasard.

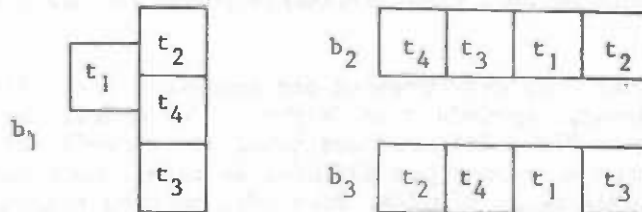
Une autre caractéristique en est le principe de répétition : tout traitement doit être appliqué au moins deux fois, sur deux placettes différentes.

On trouve ces deux caractéristiques dans des dispositifs expérimentaux très divers : dispositifs totalement aléatoires, "lattices", carrés latins, dispositifs en blocs incomplets, en "split-plots",...etc. Les manuels courants, comme celui de SNEDECOR (voir la bibliographie), traitent de la bonne utilisation et de l'analyse de ces dispositifs. DAWKINS donne des fiches de calcul pour un grand nombre d'entre eux dans son ouvrage intitulé "Statforms".

En fait, seuls deux dispositifs sont largement utilisés dans les études d'accroissement et de production : les expériences en blocs aléatoires et les expériences factorielles.

231.1 Expériences en blocs aléatoires

La figure ci-dessous donne le schéma-type d'une expérience en blocs complets aléatoires ⁽³⁾ :



⁽¹⁾ en anglais = randomized designs

⁽²⁾ en anglais = treatment

⁽³⁾ en anglais = complete randomized block experiment

Les traitements, en nombre quelconque ≥ 2 , sont désignés par t_1, t_2, \dots etc. Ils sont groupés en blocs, en nombre quelconque ≥ 2 , désignés par b_1, b_2, \dots etc, chaque bloc contenant une "répétition" ⁽¹⁾ (c'est-à-dire une réalisation) de chaque traitement.

Les blocs doivent être implantés de telle manière que les variations de type de station ou de peuplement à l'intérieur d'un bloc soient faibles devant les variations entre blocs. Il n'est pas nécessaire que les placettes appartenant à un même bloc se touchent, comme dans la figure ci-dessus, mais la distance qui peut les séparer est généralement faible devant la distance qui sépare les différents blocs. L'attribution des traitements aux placettes d'un même bloc se fait à l'aide d'une table de nombres au hasard.

Si l'un (ou plusieurs) des traitements est absent de l'un (ou plusieurs) des blocs, on se trouve devant un dispositif en blocs aléatoires incomplets. Cet état de chose peut résulter d'une décision a priori, notamment quand le nombre de traitements expérimentés est grand, ou de la perte accidentelle d'une ou plusieurs placettes. L'analyse de variance d'une expérience en blocs incomplets aléatoires est un peu plus complexe que celle d'une expérience en blocs complets, mais il n'y a pas de différences quant au traitement des résultats par régression.

Dans les études de production, les expériences en blocs aléatoires sont utiles quand les traitements expérimentés ne forment pas un continuum dimensionnel. Si, par exemple, les différents traitements portent sur l'espacement initial d'une plantation, on peut faire varier celui-ci de façon continue : un dispositif clinal, (voir § 232.2) sera alors plus approprié. Mais si les différents traitements consistent en spécifications pour l'exploitation d'une forêt dense humide, exprimées en termes de groupes d'essences et de diamètres d'exploitabilité dans chaque groupe, ils ne forment pas un continuum : on ne peut plus utiliser un dispositif clinal, et un dispositif en blocs aléatoires conviendra très bien.

231.2 Expériences factorielles

On réalise des expériences factorielles ⁽²⁾ quand chaque "traitement" est défini par une combinaison des modalités de deux ou plusieurs facteurs. Par exemple, on peut décider que l'âge à l'éclaircie et l'intensité de l'éclaircie constitueront deux facteurs distincts d'une expérience d'éclaircie.

Chaque modalité d'un facteur se combine avec chaque modalité des divers autres facteurs pour constituer autant de traitements qu'il y a de combinaisons. Ainsi, si un facteur comporte trois modalités et un autre deux modalités, on obtiendra six traitements (3×2).

Chaque traitement doit être appliqué au moins deux fois, sur deux placettes différentes. On obtient de meilleurs résultats en groupant les répétitions en blocs, car on peut ainsi prendre en compte d'éventuelles variations entre blocs.

(1) en anglais = replicate

(2) en anglais = factorial experiments

Des traitements à définition qualitative ("élagage" ou "pas d'élagage") peuvent naturellement être inclus dans une expérience factorielle sous forme de facteurs à deux modalités (présence - absence).

En peuplement homogène, les expériences factorielles sont bien adaptées aux études mettant en jeu la date et l'intensité d'éclaircie, l'espacement initial, l'élagage et l'utilisation d'engrais et d'herbicides. Il est plus difficile d'en tirer des résultats significatifs en forêt hétérogène en raison de la complexité de la définition et des effets d'une sylviculture.

232 Dispositifs systématiques

Dans les dispositifs systématiques ⁽¹⁾, la localisation des traitements n'est pas faite au hasard, mais organisée suivant un schéma systématique pour limiter l'ampleur et le coût de l'expérience. Ces dispositifs ne peuvent pas faire l'objet d'une analyse de variance, mais ils constituent un moyen très efficace d'obtenir des données pour l'ajustement d'un modèle par régression.

Ces dispositifs sont particulièrement adaptés aux expériences d'espacement ⁽²⁾ en peuplement homogène. Il faut en revanche utiliser un dispositif aléatoire chaque fois qu'il y a le moindre doute quant à la nature de la réponse aux traitements de l'expérience, par exemple dans les expériences d'utilisation d'engrais.

Dans les expériences d'espacement, on emploie surtout les deux types suivants de dispositifs systématiques :

232.1 Dispositifs mono-arbre

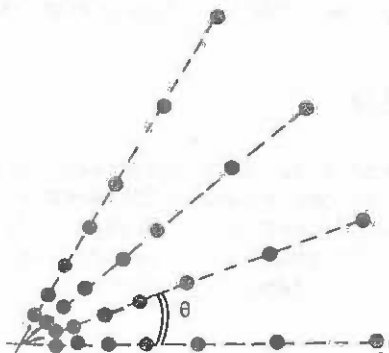
Dans un dispositif systématique d'espacement mono-arbre ⁽³⁾, l'espacement entre arbres varie de façon continue. L'éventail de NELDER ⁽⁴⁾ en est un exemple bien connu : les arbres sont plantés sur des rayons issus d'un centre, l'intervalle entre deux arbres consécutifs croissant régulièrement comme l'écartement entre les rayons (ces intervalles sont donc en progression géométrique de raison $\frac{1}{1-\theta}$, si θ est la valeur en radians de l'angle entre rayons). On obtient la disposition suivante :

(1) en anglais = systematic designs

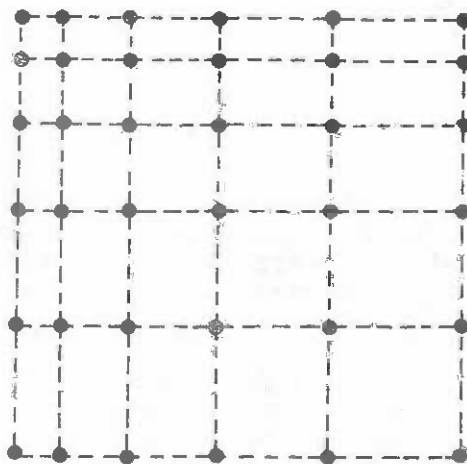
(2) en anglais = spacing experiments

(3) en anglais = single tree design.

(4) en anglais = NELDER fan



L'organisation et la réalisation de ces éventails de NELDER présentent quelque difficulté. Mais il existe d'autres dispositifs mono-arbre, de forme rectangulaire, qui procèdent du même principe et sont tout aussi efficaces ; la progression des intervalles peut y être quelconque (arithmétique, géométrique,...) . Le schéma ci-dessous illustre un exemple de dispositif rectangulaire dans lequel la distance entre lignes et entre colonnes augmente chaque fois de 0,50 m ; son avantage est de présenter toutes les configurations d'espacement en carré et en rectangle, avec en outre répétition des configurations rectangulaires. Une progression moins rapide de la distance conduirait à un dispositif plus important mais limiterait l'impact de la perte accidentelle d'un arbre.

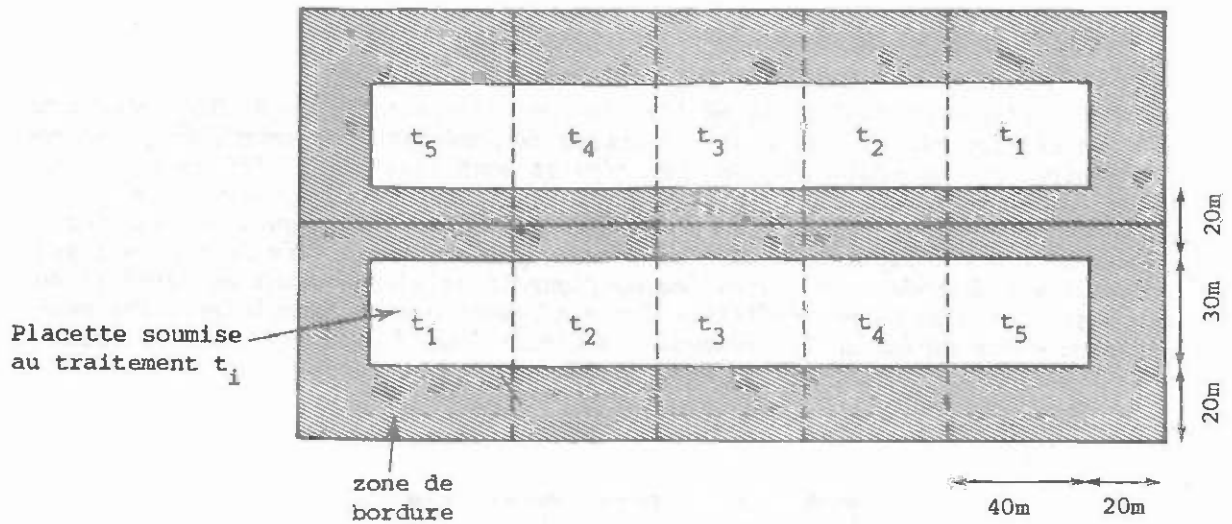


Les expériences d'espacement mono-arbre fournissent des données intéressantes sur la réponse à l'espacement, en peuplement homogène, du diamètre du tronc et du houppier d'un arbre. Mais ces données ne sont utilisables avec celles pro-

venant de placettes conventionnelles que si l'on a adopté une stratégie de modélisation individuelle des arbres. Par ailleurs, ces expériences sont très sensibles à la perte accidentelle d'arbres, qui modifie la position relative des arbres restants.

232.2 Placettes clinales

Dans un dispositif clinal ⁽¹⁾, deux placettes adjacentes se voient attribuer des traitements voisins, ce qui suppose évidemment que ces traitements puissent être classés sans ambiguïté dans un continuum. L'avantage d'un tel dispositif est de rendre les bordures ⁽²⁾ inutiles ailleurs que sur les côtés externes de l'expérience, comme le montre le schéma ci-dessous :



Dans les expériences d'espacement, il est généralement souhaitable d'avoir entre 5 et 8 traitements. Il faut toujours y incorporer (voir paragraphe 21) au moins les deux extrêmes suivants : 2×2 m ou $2,5 \times 2,5$ m selon que l'essence étudiée présente un houppier étroit (pins, eucalyptus,...) ou large (Gmelina arborea,...), et 7×7 m ou 8×8 m.

Les données tirées de placettes clinales peuvent être combinées sans problèmes avec des données venant d'autres types de placettes. D'autre part, la perte éventuelle de quelques arbres dans ce type d'expérience a moins de conséquence que dans les placettes mono-arbres.

(1) en anglais = clinal design

(2) Rappelons que la "bordure" d'une placette d'expérience reçoit le même traitement que la placette, mais ne fait pas l'objet de mesures.

24 EXEMPLES D'EXPERIMENTATION SUR LA CROISSANCE ET LA PRODUCTION

Les manières de concevoir et de réaliser des expériences fournissant des données pour les études de production sont infiniment variées. Nous allons présenter maintenant quelques exemples typiques d'expériences courantes, en discutant de leur utilité et des problèmes posés par leur réalisation et l'analyse de leurs résultats. Il faut distinguer ici les peuplements homogènes (monospécifiques équiennes) des peuplements hétérogènes (plusieurs essences et âges variés) car les techniques à utiliser y sont très différentes.

241 Peuplements homogènes

Les études menées en peuplement homogène ⁽¹⁾ sur la croissance et la production se sont concentrées sur les effets de la densité, de la fertilisation et de l'élagage. Nous ne parlerons pas ici des expériences de fertilisation et d'élagage car leur objet dépasse la prévision de la production proprement dite. En revanche, l'étude des effets de la densité du peuplement est primordiale, puisque c'est la principale variable que le forestier est capable de contrôler au cours de la vie d'un peuplement équiennne. Il y a au moins quatre façons d'étudier expérimentalement l'interaction entre la densité et la croissance d'un peuplement.

241.1 Expériences d'espacement

On peut réaliser des expériences d'espacement sous forme de dispositifs mono-arbre ou de placettes clinales, celles-ci étant sans doute plus efficaces. Il vaut mieux amener les placettes à l'espacement voulu par l'expérience par une éclaircie au cours de la deuxième ou troisième année de croissance, plutôt que de planter d'emblée à l'écartement voulu (risque d'une mauvaise reprise après la plantation). L'éclaircie qui amène à l'espacement voulu ne doit pas être faite par le haut ou par le bas, mais être rigoureusement non sélective.

La durée d'une expérience d'espacement doit couvrir la totalité de la révolution "normale" du peuplement étudié ; une fois mise en place, elle ne doit pas être éclaircie.

Il est facile d'analyser les données tirées d'une telle expérience pour construire un modèle d'accroissement de l'essence étudiée : il en sera traité en détail au chapitre 5.

241.2 Expériences d'éclaircie à surface terrière constante

Dans une expérience d'éclaircie ⁽²⁾ à surface terrière constante, on implante des placettes systématiquement ou au hasard, et on laisse croître chacune jusqu'à ce qu'elle ait atteint la surface terrière qui lui a été assignée. On l'éclaircit alors tous les ans ou tous les deux ans pour la maintenir à cette surface terrière.

Pour que ce genre d'expérience soit vraiment utile, il faut enregistrer soigneusement le nombre de tiges et la surface terrière que l'on enlève à chaque éclaircie.

L'analyse en est plus difficile que celle d'une expérience d'espacement et implique l'ajustement d'un modèle prédisant l'accroissement en surface terrière

(1) en anglais = uniform forest

(2) en anglais = thinning experiment

en fonction de la surface terrière et d'autres variables, âge et nombre de tiges sur pied par exemple.

Utilisé pour bâtir une expérience d'éclaircie dynamique, ce dispositif a l'avantage de la simplicité ; mais sa réalisation peut quand même poser des problèmes, puisqu'il faut consigner très régulièrement les observations si l'on veut en tirer des résultats vraiment utiles.

241.3 Expériences d'éclaircie avec traitements d'intensité croissante

Dans ce type d'expérience, on définit un certain nombre de traitements (souvent 4) qui diffèrent les uns des autres quant à la rotation et à l'intensité des éclaircies, et parfois quant à l'espacement initial ; ces modalités d'éclaircie peuvent être classées, de "faible" à "forte".

L'enregistrement et l'analyse des résultats de ce type d'expériences sont très semblables à ceux des expériences d'éclaircie à surface terrière constante : il faut noter le nombre et la surface terrière des tiges enlevées à chaque éclaircie. Entre autres méthodes d'analyse, la meilleure est celle qui consiste à ajuster aux résultats un modèle d'accroissement en diamètre moyen (ou en surface terrière) en fonction de la surface terrière, du nombre de tiges, de l'âge, ...etc.

Ce type d'expérience présente, sur les expériences à surface terrière constante, l'avantage d'étudier des modalités d'éclaircie réellement utilisables dans la pratique : l'expérience donne l'occasion d'évaluer leur coût, leurs effets sur la qualité du bois, les dégâts d'exploitation et les risques de chablis. En outre, les éclaircies n'y sont pas obligatoirement aussi fréquentes, ce qui simplifie la gestion du dispositif.

241.4 Expériences factorielles sur les caractéristiques des éclaircies

En peuplement homogène, un traitement sylvicole peut être analysé en un certain nombre de facteurs tels que :

- espacement initial
- âge à la première éclaircie
- proportion du matériel sur pied enlevée à chaque éclaircie
- rotation à respecter (intervalle entre deux éclaircies).

On peut évidemment choisir des facteurs différents, tels que la surface terrière au moment de chaque éclaircie, ou la hauteur au lieu de l'âge, ...etc.

Dans la logique de cette analyse, on peut concevoir l'expérience d'éclaircie comme une expérience factorielle, avec plusieurs modalités pour chacun des facteurs de l'éclaircie. Il en résultera un dispositif de grande envergure avec un nombre considérable de traitements et de placettes. Mais l'avantage en sera, lors de l'analyse des résultats, la possibilité de séparer les effets respectifs de l'âge de l'éclaircie, de son intensité, ...etc.

Il est évident qu'un dispositif aussi important et complexe demande une gestion très attentive, tandis que le caractère factoriel de l'expérience peut être remis en question par un quelconque accident (incendie, épidémie, ...).

242 Peuplements hétérogènes

Faire une expérience en forêt hétérogène ⁽¹⁾, c'est généralement y introduire une perturbation contrôlée, de façon à pouvoir ajuster par la suite un modèle reliant l'accroissement des arbres aux divers paramètres du peuplement soumis à ce traitement.

La cause principale des problèmes que l'on y rencontre est que les traitements sont souvent définis en des termes qui n'ont que peu de rapport avec les paramètres du peuplement qui régissent sa croissance. Pour illustrer cela, on peut citer un exemple typique, avec définition de quatre traitements :

- 1 - Exploitation de toutes les essences commercialisables à partir de 30 cm de diamètre.
- 2 - Comme 1, mais avec empoisonnement ou annellation de toutes les essences non commercialisables à partir de 30 cm de diamètre.
- 3 - Comme 2, mais en tuant toutes les tiges non commercialisables à partir de 10 cm de diamètre.
- 4 - Comme 3, mais en enlevant toutes les tiges d'essences commercialisables à partir de 20 cm de diamètre.

Les effets sur le peuplement de traitements ainsi définis dépendront entièrement de l'état du peuplement avant l'exploitation, de sa distribution par essences et par catégories de diamètre. Il se peut même que le traitement le plus énergique y ait le moins d'effet.

Le choix, séduisant à première vue, d'un sous-échantillon d'arbres fondé sur des critères de commercialisation pour les mesures d'accroissement a conduit à des échecs, car ces critères ont une fâcheuse tendance à changer entre deux mesures de la placette. Par ailleurs, les méthodes utilisées pour mesurer les accroissements ont rarement résolu convenablement le problème posé par le développement des contreforts.

Actuellement, les recherches sur la croissance et la production en forêt tropicale hétérogène sont en plein développement. Les recommandations qui suivent ne sont donc là que pour éviter certaines erreurs du passé, non pour imposer aux travaux en cours une réglementation inutilement rigide.

242.1 Dispositifs en blocs aléatoires

Dans les expériences en forêt hétérogène, nous recommandons de toujours adopter un dispositif aléatoire, avec répétition ; ces répétitions doivent être groupées en blocs aussi homogènes que possible en ce qui concerne les caractéristiques de la station, de la composition du peuplement et des exploitations passées. Tout cela implique une étude préalable très complète de la zone d'expérience.

⁽¹⁾ en anglais = mixed forest

242.2 Définition du traitement sylvicole

Les traitements doivent être définis en termes de surface terrière totale des arbres de diamètre supérieur à 10 cm restant après exploitation et/ou empoisonnement ou annellation.

On peut les définir autrement que par la surface terrière, mais ce sera toujours en termes de peuplement restant sur pied (et non de matériel à enlever), et toujours indépendamment de tout critère économique.

L'expérience comportera toujours deux traitements extrêmes : l'un consistera en une absence de traitement (placettes témoins), l'autre en un traitement très sévère, ne laissant par exemple aucune tige de plus de 10 cm de diamètre.

242.3 Caractéristiques des placettes - Mesures à faire

Quand les traitements mettent en jeu l'exploitation, les placettes doivent être de grande dimension : généralement 200 x 200 m (4 ha) avec une bordure de protection large de 100 m soumise au même traitement. L'emplacement de tous les gros arbres (de diamètre > 30 cm, par exemple) de la placette doit être cartographié. Une subdivision de la placette en carrés de 20 x 20 m est souhaitable et peut permettre d'inclure les effets de la compétition locale dans les modèles d'accroissement, de régénération et de mortalité.

On peut faire un inventaire détaillé de la régénération sur un sous-échantillon systématique de carrés.

Les méthodes de mesure des accroissements sont exposées au chapitre 3.

CHAPITRE 3

RECUEIL DES DONNEES
ET PREMIER TRAITEMENT

3 RECUEIL DES DONNEES ET PREMIER TRAITEMENT

31 DELIMITATION DES PLACETTES D'ECHANTILLONNAGE

311 Localisation

Les placettes permanentes d'échantillonnage doivent être localisées avec précision sur les cartes forestières, et leur position exacte dans la forêt déterminée à l'aide du ruban et de la boussole. Il est bon, en outre, de poser sur les routes forestières proches des repères de pierre ou de béton indiquant la direction et la distance de la placette.

Les placettes temporaires doivent aussi être localisées sur les plans, mais avec une précision qui peut être moindre.

Les principes présidant au choix de l'emplacement des placettes ont été définis au § 22.

312 Identification des placettes permanentes sur le terrain

Les placettes permanentes doivent être repérées sur le terrain de façon durable. Une placette circulaire sera repérée par un poteau de bois, de ciment ou de métal planté en son centre et portant son numéro d'identification. Il est bon de signaler en outre ce centre par l'intersection de deux fossés de 50 cm de profondeur et de 2,50 m de longueur environ : on a ainsi sur le sol une marque permanente qui peut se révéler bien utile en cas de disparition du poteau central.

De même, les quatre coins d'une placette rectangulaire seront repérés par des poteaux, dont l'un portera le numéro d'identification de la placette, et des fossés se coupant en chacun de ces points.

Que la placette soit circulaire ou rectangulaire, son numéro d'identification sera en outre peint de façon très visible sur un arbre proche du poteau sur lequel il figure déjà.

Si l'on subdivise une placette rectangulaire en petits carrés, on peut planter à chacun de leurs sommets un piquet, plus petit que les poteaux des coins pour éviter toute confusion ; mais on peut aussi redélimiter ces carrés lors de chaque mesure.

313 Limites de la placette et arbres-limites

La plupart des arbres se trouveront manifestement à l'intérieur ou à l'extérieur de la placette, mais certains seront à cheval sur sa limite : ils appartiendront alors à la placette si le centre supposé du fût est à l'intérieur de la limite.

La limite d'une placette circulaire est déterminée au moyen d'une corde ou cordelette tendue horizontalement depuis le centre. Il faut veiller à ce que la corde n'ait aucune élasticité (certaines cordelettes de nylon sont très élastiques), et ne s'enroule pas autour du poteau central quand on fait le tour de la placette. Le tableau suivant indique le rayon à donner à une placette circulaire pour que sa surface ait une valeur donnée.

Surface de la placette (ha)	Rayon (m)
0,04	11,28
0,05	12,62
0,08	15,96
0,10	17,84

Pour les petites placettes rectangulaires, la limite peut être déterminée par la ligne de visée entre les poteaux de deux coins contigus. Lorsqu'elles sont plus grandes et que l'on est gêné par la végétation, on est obligé d'intercaler des repères supplémentaires le long de l'orientation joignant un coin au suivant.

Il peut être judicieux de couper la végétation le long de la limite, à condition de ne pas détruire la régénération, qui constitue une partie du peuplement forestier.

314 Identification des arbres

Dans la mesure du possible, les arbres d'une placette permanente doivent être individuellement identifiés. Il y a deux moyens de le faire :

- 1 - En peignant un numéro d'identification sur chaque arbre.
- 2 - En y clouant une plaquette d'aluminium portant un numéro en relief.

Si on peint le numéro d'identification sur l'arbre, on repérera aussi par un trait de peinture autour de l'arbre l'emplacement de la mesure du diamètre.

Si on utilise une plaquette métallique, on la clouera sur l'arbre à une hauteur déterminée (généralement 50 cm) au-dessus de l'emplacement de la mesure du diamètre, pour pouvoir facilement repérer cet emplacement par la suite. Rappelons que le niveau normal de référence pour la mesure du diamètre est 1,30 m au-dessus du sol, sauf pour les arbres à contreforts ; si le niveau du sol change au cours de la vie de l'arbre, l'emplacement de la mesure du diamètre ne doit pas changer : c'est sa distance au sol qui deviendra différente de 1,30 m.

Il faut aussi clouer une plaquette d'aluminium au pied de l'arbre, tout près du sol, pour faciliter l'identification des souches d'arbres abattus et se prémunir contre la perte éventuelle de la plaquette supérieure.

On se gardera d'utiliser la peinture comme seul moyen d'identification sur des essences qui perdent leur écorce (comme beaucoup d'Eucalyptus).

On doit remplacer ou renouveler les marques chaque fois que l'on constate, à l'occasion d'une mesure, leur disparition ou leur usure.

315 Cartographie des arbres de la placette

Dans la mesure du possible, on cartographiera l'emplacement de tous les arbres d'une placette permanente au moment de son implantation. Si la placette est circulaire, on notera la distance et l'azimut de chaque arbre par rapport au

centre. Si la placette est rectangulaire, on la subdivisera en carrés ne dépassant pas 10 m de côté, puis on mesurera la distance de chaque arbre à deux côtés du carré où il se trouve pour avoir ses coordonnées dans ce carré.

Cette cartographie permet de résoudre les confusions qui se produisent fréquemment sur l'identité des arbres, et d'envisager une analyse fine des phénomènes de croissance sur la placette.

316 Identification des arbres passés à la futaie

Dans les forêts naturelles, il faut attribuer aux arbres passés à la futaie depuis la mesure précédente un numéro d'identification, tout en notant leurs coordonnées et leur numéro de carré. Ce numéro d'identification ne doit avoir été attribué précédemment à aucun arbre de la placette, même mort ou exploité, sous peine de la plus grande confusion au moment de l'exploitation des données.

32 FICHES DE MESURES POUR LES PLACETTES D'ECHANTILLONNAGE ET PREMIER TRAITEMENT DES DONNEES

321 Peuplements homogènes

Dans les peuplements équiennes monospécifiques (souvent obtenus par plantation), les placettes d'échantillonnage feront l'objet des mesures suivantes :

- 1 Diamètre à 1,30 m sur écorce de chaque arbre, à l'aide d'un ruban gradué en π cm.
- 2 Hauteur totale de 8 arbres constituant un échantillon systématique, et des arbres dominants non compris dans cet échantillon.

La hauteur dominante est la moyenne des hauteurs des 100 plus gros arbres à l'hectare. On espère en avoir une estimation valable, sur une placette de surface s , en mesurant la hauteur des 100 \times s plus gros arbres de la placette (avec s en ha) : si la placette fait 0,04 ha, on estimera la hauteur dominante du peuplement à partir de la hauteur des 4 plus gros arbres. Les arbres dont la cime est cassée ou sérieusement endommagée seront rejetés des échantillons de mesure de la hauteur.

Chaque arbre peut faire l'objet d'observations complémentaires : on peut noter la présence de maladies, les arbres mourants, les dégâts dus au vent ou aux insectes, les arbres marqués en éclaircie. Ces observations doivent être notées selon un code strictement standardisé, dont un exemple est suggéré ci-dessous :

<u>Code</u>	<u>Description</u>
rien	Arbre sain, non endommagé, non marqué en éclaircie, à une seule tige sans défaut.
A-	Arbre abattu (il peut être mesuré au sol).
C	Arbre cassé par le vent.

<u>Code</u>	<u>Description</u>
D	Dégâts d'animaux.
E	Arbre en cours d'élimination par la concurrence.
M	Arbre marqué en éclaircie (mais toujours sur pied).
P	Arbre penché ou renversé par le vent.
S	Problème sanitaire.
T	Tige ou flèche double ou multiple.
X	Arbre mort.

On peut noter après ces lettres un chiffre indiquant le degré de gravité du problème, par exemple selon l'échelle suivante :

- 1 Présence (de dégâts ou de maladie) mais sans incidence sérieuse.
- 2 Plus grave, susceptible de réduire sérieusement la croissance ou les possibilités d'utilisation.
- 3 Très grave, de nature à tuer l'arbre ou à le rendre inutilisable.

On peut encore ajouter d'autres codes pour traiter des problèmes particuliers. Ce qui importe pour que ce système soit utile, c'est de le suivre fidèlement sur de nombreuses années sans omission ni modification.

L'imprimé 3.1 est un exemple de fiche de mesure pour ce type de placette. Elle est conçue spécialement pour faciliter le traitement automatique des données. Elle doit être remplie directement sur le terrain, au crayon à mine tendre et en utilisant une gomme pour effacer les erreurs.

Le premier traitement consiste à calculer les variables suivantes pour chaque placette :

- 1 Nombre de tiges par hectare (N), obtenu en divisant le nombre total d'arbres vivants sur la placette par la surface totale de la placette ; si l'on se réfère à l'exemple de la fiche 3.1, ceci donne :

$$N = 9/0,04 = 225 \text{ tiges/ha.}$$

- 2 Diamètre de l'arbre de surface terrière moyenne (D_g) ; c'est la moyenne quadratique des diamètres, obtenue en additionnant les carrés des diamètres, en divisant par le nombre d'arbres de la placette et en prenant la racine carrée du résultat. Pour l'exemple de la fiche 3.1 :

$$\Sigma d^2 = 16\ 237$$

$$D_g = \sqrt{16\ 237/9} = 42,5 \text{ cm}$$

IMPRIME 3.1

: FICHE DE MESURE D'UNE PLACETTE-ECHANTILLON - PLANTATIONS

Page : 1
 Nombre de pages remplies : 1

Forêt	Série	Parcelle	Placette n°	Essence
103	002	12A	2397	CL
Surface de la placette		Pente (en grades)	Mois	Année
0400		6	07	79
Date de la mesure:				2

Nombre de cartes (ne pas remplir)

	Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	Codes		
1ère carte	11	39.4	28.6			
	16	43.1	32.1			
	13	36.4	.	C3	X	
	15	48.5	29.6			
	19	46.4	30.8			
2ème carte	21	41.2	31.3			
	24	38.1	34.2			
	4	42.8	33.8			
	9	44.9	33.1			
3ème carte		.	.			
		.	.			
		.	.			
		.	.			
		.	.			
4ème carte		.	.			
		.	.			
		.	.			
		.	.			
		.	.			
5ème carte		.	.			
		.	.			
		.	.			
		.	.			
		.	.			

NOTES Arbre 13 frappé par la foudre ? Cassé à mi-hauteur et mort

RESPONSABLE DE LA MESURE : P. ALDER DATE : 7.7.78

FAIRE LES CALCULS DE HAUTEUR AU VERSO.

- 3 Surface terrière du peuplement (G), obtenue en multipliant par $\pi/40\ 000$ la Σd^2 ci-dessus et en divisant le résultat par la surface de la placette ; elle est alors exprimée en m²/ha :

$$G = (16\ 237 \times \pi/40\ 000)/0,04 \\ = 31,88 \text{ m}^2/\text{ha}$$

- 4 Hauteur dominante du peuplement (H_0) ; c'est la moyenne des hauteurs des arbres de l'échantillon "dominant", c'est-à-dire ici des 4 plus gros de la placette :

$$H_0 = (29,6 + 30,8 + 33,1 + 32,1)/4 \\ = 31,4 \text{ m}$$

- 5 Hauteur moyenne du peuplement (H) ; c'est la moyenne des hauteurs des 8 arbres de l'échantillon "systématique" (si la placette porte moins de 8 arbres, l'échantillon systématique comprend tous les arbres de la placette). Pour l'exemple de la fiche 3.1 :

$$H = 31,7 \text{ m}$$

- 6 Volume du peuplement (V) ; on le calcule généralement à partir d'un tarif de cubage individuel à deux entrées, le diamètre à 1,30 m et la hauteur totale. Il y a deux méthodes :

- (i) Calculer, à l'aide du tarif, le volume de l'arbre de diamètre D et de hauteur H, et multiplier le chiffre obtenu par le nombre de tiges/ha N pour obtenir le volume en m³/ha.
- (ii) Calculer le volume v de chaque arbre à partir de son diamètre à 1,30 m d et de sa hauteur totale h, additionner ces volumes et diviser par la surface de la placette pour obtenir le volume à l'hectare. Si on ne connaît pas la hauteur h de tous les arbres, on l'estimera à partir d'une régression hauteur/diamètre construite sur l'échantillon systématique (voir § 35).

Ces deux méthodes introduisent une erreur dans l'estimation du volume. La première peut présenter un biais résultant de la distribution des diamètres et des hauteurs, la seconde une erreur aléatoire due aux mesures de hauteur. La seconde méthode est préférable pour sa précision, surtout si l'on estime le volume jusqu'à une découpe marchande fixée en diamètre ⁽¹⁾.

⁽¹⁾ en anglais = merchantable top diameter limit

On peut calculer plusieurs volumes en utilisant différents tarifs de cubage, par exemple :

- V_0 = volume sur écorce ⁽⁸⁾ de la tige totale (jusqu'au sommet de l'arbre)
 - V_{0s} = volume sous écorce ⁽⁹⁾ de la tige totale (jusqu'au sommet de l'arbre)
 - V_7 = volume sur écorce de la tige jusqu'à la découpé 7 cm de diamètre (volume de la tige bois-fort)
 - V_{tot} = volume sur écorce de l'arbre total (tige et branches jusqu'au bout).
- ... etc...

Les divers symboles et unités utilisés pour les variables fondamentales sont récapitulés dans le tableau 3.1.

Tableau 3.1

Symboles utilisés pour les variables fondamentales du peuplement

Symbole	Description	Unité
A	Age du peuplement depuis sa plantation ⁽¹⁾	Année
D_g	Diamètre de l'arbre de surface terrière moyenne ⁽²⁾	cm
G	Surface terrière du peuplement/ha ⁽³⁾	m ² /ha
H	Hauteur moyenne du peuplement ⁽⁴⁾	m
H_0	Hauteur dominante du peuplement ⁽⁵⁾ , définie comme étant la hauteur moyenne des 100 plus gros à l'hectare	m
N	Nombre d'arbres vivants/ha ⁽⁶⁾	Arbres/ha
V	Volume du peuplement/ha ⁽⁷⁾ Le volume dont il s'agit doit absolument être précisé ; sa nature peut être rappelée par un indice (voir le texte).	m ³ /ha

On peut repérer les mesures faites immédiatement après éclaircie par une apostrophe (') ; exemple : N' = nombre de tiges/ha laissées sur pied après éclaircie. On signalera de même le matériel enlevé lors d'une éclaircie par un e en indice ; exemple : G_e = surface terrière/ha enlevée en éclaircie.

- ⁽¹⁾ en anglais = stand age from planting
- ⁽²⁾ en anglais = diameter of the mean basal area tree
- ⁽³⁾ en anglais = stand basal area
- ⁽⁴⁾ en anglais = stand mean height
- ⁽⁵⁾ en anglais = stand dominant height, ou top height
- ⁽⁶⁾ en anglais = stocking/ha ou number of live trees/ha
- ⁽⁷⁾ en anglais = stand volume/ha
- ⁽⁸⁾ en anglais = over bark
- ⁽⁹⁾ en anglais = under bark

322 Forêts hétérogènes

En forêt hétérogène, le principal résultat du premier traitement des données d'une placette doit être un tableau du peuplement donnant le nombre de tiges par catégorie de grosseur et par essence ou groupe d'essences. Une fois ce tableau du peuplement construit, on pourra en tirer facilement diverses caractéristiques du peuplement sur pied : il est facile d'obtenir, par exemple, la surface terrière des arbres d'un ensemble donné d'essences dépassant une grosseur donnée ; il sera tout aussi facile, en cas de besoin, de recalculer cette caractéristique pour un autre groupe d'essences et une autre limite de grosseur.

On peut naturellement construire directement le tableau du peuplement en ne relevant, au moment de la mesure, que les nombres de tiges par catégories de grosseur et par groupes d'essences. Mais il vaut mieux ne pas procéder ainsi, sauf peut-être lors d'inventaires rapides pour une pré-évaluation de la ressource : il faut relever les dimensions et l'essence de chaque arbre, même si l'on a l'intention de résumer par la suite les données. Ce relevé est particulièrement important sur les placettes permanentes puisque c'est lui qui permettra de calculer l'accroissement de chaque arbre.

En forêt hétérogène, les fiches de mesure peuvent prendre des formes très variées, selon les caractéristiques que l'on veut y noter. Voici quelques situations typiques :

- a) Forêt tropicale hétérogène de montagne ⁽¹⁾ (quelques essences de lumière d'âges variés). On peut utiliser l'imprimé 3.1 en réservant la première colonne des "codes" à l'essence, avec un code à deux chiffres ou deux lettres. On doit toujours mesurer les hauteurs sur un échantillon systématique pour permettre la construction d'une courbe hauteur/diamètre.
- b) Forêt dense humide ⁽²⁾ (grand nombre d'essences, âges variés, lianes, épiphytes, contreforts). On y installera normalement de grandes placettes, de plus d'un hectare, subdivisées en carrés de 10 × 10 m. On ne peut généralement pas mesurer les hauteurs, mais on peut classer les arbres selon la forme du houppier et sa position dans le couvert. Les essences qu'on peut y rencontrer sont plusieurs centaines. Enfin, les arbres à contreforts font l'objet de deux mesures de diamètres de référence (voir § 342).
- c) Forêt claire sub-humide ⁽³⁾ (forêt de Miombo en Afrique orientale, par exemple). Ici, la forme et la longueur de la grume commercialisable sont des caractéristiques importantes. Le degré d'ouverture de ces forêts permet les mesures de hauteur, mais on ne peut espérer une précision meilleure que le mètre, à cause de la forme diffuse du houppier. Les essences que l'on peut y rencontrer dépassent la centaine.

(1) en anglais = mixed tropical montane forest

(2) en anglais = tropical rainforest

(3) en anglais = sub-humid woodland

- d Forêt claire en zone aride ⁽¹⁾. Les arbres y étant le plus souvent à tige multiple, l'essence et la hauteur sont les deux seules caractéristiques importantes. Les hauteurs doivent être mesurées au décimètre près.

La figure 3.1 présente des formats d'enregistrement des mesures adaptés à ces différents cas, le premier ne faisant que reprendre l'imprimé 3.1 sous une autre présentation (un "enregistrement" correspond à une carte perforée à 80 colonnes, moyen d'entrée des données en ordinateur qui est encore le plus courant).

323 Première mesure des placettes permanentes

Lorsqu'on mesure pour la première fois une placette permanente, lors de son installation, il faut noter les informations suivantes :

- 1 La surface exacte de la placette en projection horizontale.
- 2 Des informations sur la station dont latitude, longitude, exposition, altitude, pente, situation topographique, histoire de la forêt et ancienne occupation du sol.
- 3 Des informations météorologiques fournies par la station météo la plus proche avec, pour chaque mois, la moyenne des précipitations, des minima et des maxima de température.
- 4 L'emplacement de chacun des arbres de la placette. Sur une placette circulaire, il sera défini par la distance et l'azimut de l'arbre par rapport au centre de la placette. Sur une placette rectangulaire, on peut le définir par les coordonnées X et Y (en décimètres) à partir du coin le plus au Sud-Ouest de la placette, en prenant comme axes les deux côtés de la placette issus de ce sommet (voir aussi § 315).
- 5 Des informations sur le profil pédologique, comprenant :
 - . couleur
 - . texture
 - . pH
 - . résultats de l'analyse pour N, P, K, Ca, Mg
 - . profondeur
 - . densité apparente,

pour chacun des horizons du sol discernables. On doit faire sur chaque placette deux fossés ou deux trous à la tarière pédologique, et même plus sur les grandes placettes qui risquent d'être très hétérogènes. Les techniques d'analyse peuvent varier quelque peu selon les méthodes en cours dans les différents pays et les conditions particulières que l'on y rencontre : elles doivent être définies en collaboration avec les pédologues. Mais, une fois que l'on a défini ces techniques d'analyse, seules des raisons très importantes doivent conduire à en changer ; on devra alors, au moment de la modification, analyser une série d'échantillons (au moins 20 ou 30) suivant l'ancienne méthode et suivant la nouvelle pour déterminer par régression la relation liant les résultats des deux méthodes : c'est nécessaire pour pouvoir ultérieurement faire des comparaisons.

⁽¹⁾ en anglais = arid zone woodland

Plantation ou autre peuplement équiennne monospécifique.

1 ^{er} ARBRE				2 ^{ème} ARBRE				3 ^{ème} ARBRE				4 ^{ème} ARBRE				5 ^{ème} ARBRE			
Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	CODES	Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	CODES	Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	CODES	Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	CODES	Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	CODES

Forêt dense humide avec arbres à contreforts.

CARRÉ		1 ^{er} ARBRE				2 ^{ème} ARBRE				3 ^{ème} ARBRE										
Arbre n°	Code essence	Hauteur	Diamètre de référence inférieur	Diamètre de référence supérieur	Hauteur	Code essence	Diamètre de référence inférieur	Diamètre de référence supérieur	Hauteur	Code essence	Diamètre de référence inférieur	Diamètre de référence supérieur	Hauteur	Code essence	Diamètre de référence inférieur	Diamètre de référence supérieur	Hauteur	Code essence	Diamètre de référence inférieur	Diamètre de référence supérieur

Forêt sub-humide (format pour placettes temporaires : arbres non numérotés).

1 ^{er} ARBRE				2 ^{ème} ARBRE				3 ^{ème} ARBRE				4 ^{ème} ARBRE				5 ^{ème} ARBRE					
Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence	Diamètre à 1,30 m sur écorce	Hauteur	Code essence

Forêt de zone aride avec arbres à tige multiple (format pour placettes temporaires : arbres non numérotés).

1 ^{er} ARBRE		2 ^{ème} ARBRE		3 ^{ème} ARBRE		4 ^{ème} ARBRE		5 ^{ème} ARBRE		6 ^{ème} ARBRE		7 ^{ème} ARBRE		8 ^{ème} ARBRE		9 ^{ème} ARBRE		10 ^{ème} ARBRE	
Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur	Code essence	Hauteur

Figure 3.1

Formats d'enregistrement pour la mesure des arbres des placettes d'échantillonnage dans différents types de forêt.

33 ANALYSE DE TIGE

L'analyse de tige ⁽¹⁾ vise à reconstituer l'histoire de la croissance d'un arbre. Elle consiste à :

- a) abattre l'arbre,
- b) découper des rondelles dans le tronc à intervalles de 2 m environ,
- c) compter les cernes ⁽²⁾ et mesurer leur diamètre sur chaque rondelle.

Elle n'est applicable que sous les climats qui présentent des saisons assez contrastées et sur les essences à cernes bien visibles.

L'analyse de tige peut apporter beaucoup d'information sur la dynamique d'un peuplement, mais c'est surtout la reconstitution de la croissance en hauteur dominante d'un peuplement homogène qui nous intéresse ici.

Pour reconstituer la croissance en hauteur d'un arbre, il suffit de noter le nombre de cernes et la hauteur exacte aux différents niveaux mesurés. Pour les essences et les climats produisant des cernes bien nets, cette observation peut se faire sur le terrain en tronçonnant simplement la tige abattue aux emplacements des mesures.

Quand les cernes ne sont pas assez nets, il faut découper des rondelles et noter immédiatement sur la face supérieure de chacune à quelle hauteur elle a été prélevée. Elles sont ensuite envoyées au laboratoire, en évitant tout retrait, et mesurées par l'une des deux méthodes suivantes :

- 1 Rabotage de la rondelle et comptage des cernes selon deux rayons à l'aide d'un microscope fixé sur un montage à crémaillère équipé d'un vernier.
- 2 Analyse densitométrique aux rayons X de deux échantillons radiaux, pris selon deux rayons perpendiculaires.

Quelle que soit la méthode choisie, il convient de vérifier l'authenticité des cernes douteux (faux cernes ou cernes très peu marqués) en les comparant aux enregistrements climatiques. L'analyse densitométrique aux rayons X, qui fournit une valeur numérique de l'épaisseur et de la densité du cerne, permet la recherche de corrélations avec les variables climatiques.

On détermine l'âge auquel l'arbre a atteint une hauteur donnée en soustrayant le nombre de cernes à cette hauteur du nombre de cernes à la base, en supposant que ce dernier correspond à l'âge de l'arbre depuis la plantation ; on peut ainsi tracer la courbe hauteur/âge de l'arbre, qui reconstitue l'histoire de sa croissance en hauteur.

Si l'arbre ainsi analysé est un dominant bien conformé, on peut espérer que sa courbe hauteur/âge est proche de la courbe hauteur dominante/âge qu'aurait fournie l'observation d'une placette permanente installée dans le même peuplement. Il est toutefois prudent, vu la variabilité individuelle des croissances en hauteur, d'analyser ainsi celle de plusieurs arbres dominants du peuplement, choisis selon une procédure bien objective.

⁽¹⁾ en anglais = stem analysis

⁽²⁾ en anglais = growth ring

34 METHODES DE MESURE DE L'ACCROISSEMENT DES ARBRES

341 Principe de la mesure

Sur une placette permanente, on estime l'accroissement annuel en diamètre (1) d'un arbre en divisant la différence entre deux mesures successives du diamètre par l'intervalle de temps séparant ces mesures. Ce procédé ne peut donner de résultats valables que si :

- 1 Chaque arbre porte une identification claire et sans équivoque sur la placette.
- 2 L'emplacement de la mesure du diamètre est localisé avec précision sur l'arbre ; on peut employer deux méthodes :
 - (i) peindre un trait tout autour de l'arbre, au niveau 1,30 m ; mais ceci peut s'avérer trop visible dans certains cas,
 - (ii) planter un clou à une distance précise (par exemple 50 cm) au-dessus de l'emplacement de la mesure ; celui-ci sera alors repéré par rapport au clou, qui peut également servir à fixer une plaquette d'identification de l'arbre.

342 Mesures successives sur les arbres à contreforts

Sur les arbres à contreforts (2), on a l'habitude de mesurer le diamètre à un niveau de référence situé environ 1 mètre au-dessus de la fin des contreforts. Ceux-ci se développant entre deux mesures, on est alors amené à déplacer de temps à autre le niveau de référence, ce qui interdit l'estimation de l'accroissement par comparaison de mesures successives.

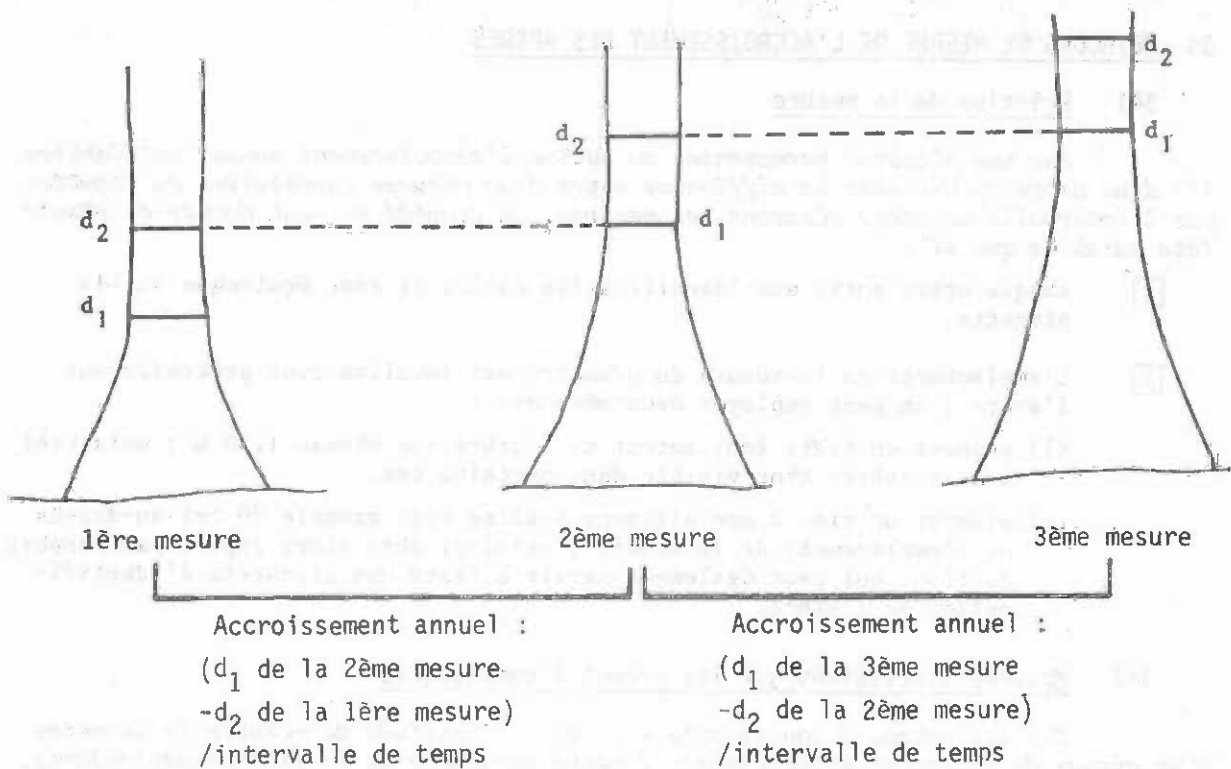
On peut envisager deux façons de résoudre cette difficulté :

- la première consiste à utiliser deux niveaux de référence à chaque mesure. Le moment venu de déplacer le niveau inférieur, on le remplacera par le niveau supérieur et on définira au-dessus un nouveau niveau supérieur : on aura ainsi au moins un diamètre de référence commun à deux mesures successives pour estimer l'accroissement.

On peut recommander d'espacer de 1,50 m les deux niveaux, et de placer le niveau inférieur 1 m au-dessus de la fin des contreforts. Le schéma ci-après illustre cette méthode.

(1) en anglais = annual diameter increment

(2) en anglais = buttresses



Ces niveaux de référence doivent être repérés par un trait de peinture ou des clous. Il faut absolument se servir d'une échelle et prendre les mesures de diamètre au ruban, avec toutes les précautions d'usage. Les instruments optiques du genre relascope ne sont pas assez précis pour l'estimation des accroissements.

- la deuxième façon de résoudre le problème consiste à utiliser des rubans dendrométriques (voir ci-dessous) et à faire les mesures d'accroissement sur des intervalles de temps assez courts (1 ou 2 ans) pour pouvoir négliger l'influence du développement des contreforts sur le diamètre de référence.

343 Rubans dendrométriques

Il est possible de fabriquer sur place et à bon marché des rubans dendrométriques ⁽¹⁾, très utiles dans les expériences de recherche intensive (expériences d'éclaircie, par exemple).

Sans vernier, ils sont précis au demi-millimètre près, ce qui convient très bien à l'estimation en un an de l'accroissement en diamètre des essences tropicales. Avec un vernier, ils atteignent une précision sur le diamètre de 0,05 mm, ce qui est suffisant pour mettre en évidence des réponses physiologiques et des fluctuations saisonnières de la croissance. En comparaison, les mesures de circonférence usuelles permettent d'obtenir l'accroissement en diamètre avec une précision qui atteint au mieux le demi-centimètre sur les petits arbres, et bien pire pour les gros arbres.

⁽¹⁾ en anglais = girth band

Voici comment procéder pour construire ces rubans :

- 1 Fournitures : du ruban en acier inoxydable de 1 cm de large, une bonne longueur de ressort à boudin de 4 mm, des gabarits pour la graduation principale et pour celle du vernier, de la peinture céramique noire mate, un emporte-pièce pour les encoches de fixation du ressort, et un four électrique ou à gaz.
- 2 Enrouler sur un bidon vide plusieurs rubans découpés à la bonne longueur, et les peindre sur une face avec la peinture noire. Mettre au four le bidon, avec ses rubans, pour cuire la peinture.
- 3 Incrire le point zéro à l'un des bouts du ruban en rayant la peinture à l'aide d'une pointe de métal. Si le ruban doit comporter un vernier, sa graduation commence aussi en ce point, la position des dix traits suivants étant définie par le gabarit du vernier (10 unités du vernier = 9 unités de la graduation principale).
- 4 Il faut connaître la circonférence approximative de l'arbre auquel est destiné le ruban pour le couper à la bonne longueur et tracer au bon endroit, sur une dizaine de cm (correspondant à 3 cm en diamètre), la graduation principale. Tracer cette graduation sur le ruban à l'aide du gabarit approprié.
- 5 Percer les encoches de fixation du ressort, l'une en-deçà du point zéro (et du vernier), l'autre dans la graduation principale. Fixer une extrémité du ressort dans l'une des encoches ; placer le ruban autour de l'arbre, dans un plan bien perpendiculaire à l'axe du tronc au niveau de référence, en prenant soin d'enlever les morceaux d'écorce qui se détachent, puis fixer à l'autre encoche l'extrémité libre du ressort ; le ressort doit être coupé à une longueur qui assure une bonne tension du ruban.

L'accroissement est mesuré par le déplacement de la graduation entre les deux mesures. Si la graduation principale a été tracée en π mm, on obtient directement l'accroissement en diamètre. Si l'on a tracé le début de la graduation principale à une distance connue du point zéro, le ruban donnera en outre la valeur du diamètre lors de chaque mesure ; sinon, il faudra mesurer le diamètre à l'aide d'un ruban normal, appliqué immédiatement au-dessus ou au-dessous du précédent.

On peut se procurer les gabarits pour graver les rubans chez des fournisseurs d'instruments forestiers, ou les fabriquer localement dans un atelier bien équipé.

Il est possible d'acheter directement divers types de rubans dendrométriques sophistiqués, dont certains équipés pour la télémessure. Leur prix élevé les réserve aux formes de recherche les plus intensives dans lesquelles la zone expérimentale peut faire l'objet d'une surveillance attentive contre les dégâts dus aux hommes ou aux animaux.

344 Mesures sur les cernes

Quand les cernes sont visibles, on peut les utiliser pour estimer l'accroissement. La méthode la plus sûre consiste à observer la section de la tige au niveau 1,30 m sur des arbres abattus.

Il faut mesurer l'épaisseur des trois ou quatre derniers cernes annuels sur deux diamètres perpendiculaires de la section. Si la section est elliptique, on fera ces mesures sur le petit axe et le grand axe de l'ellipse. On obtient ainsi l'accroissement périodique sous-écorce ; pour le convertir en accroissement sur écorce cohérent avec les mesures normales de diamètre à 1,30 m, on établira et utilisera une régression reliant la mesure directe du diamètre sous écorce à la mesure au ruban du diamètre sur écorce.

Il est possible d'estimer l'accroissement en diamètre d'un arbre à partir de carottes prélevées à l'aide d'une tarière de Pressler, mais cette méthode peut donner lieu à des erreurs, surtout avec les essences à bois tendre : la carotte obtenue peut être déformée de diverses façons par la libération des contraintes internes du bois, ou ne pas être exactement perpendiculaire aux cernes.

Vu la difficulté qu'il y a à compter les cernes en zone tropicale, imposant souvent le recours au microscope calibré ou à l'analyse densitométrique aux rayons X, il est presque obligatoire de disposer de sections complètes : l'usage de la tarière de Pressler y est donc rarement possible.

35. ESTIMATION INDIRECTE DE LA HAUTEUR DES ARBRES ET DE LA HAUTEUR DOMINANTE

Les mesures de hauteur étant relativement longues, il vaut mieux se limiter à la mesure de la hauteur de 8 à 10 arbres par placette. Si l'on a besoin de la hauteur de tous les arbres de la placette, on peut construire à partir de l'échantillon mesuré une courbe hauteur/diamètre de la forme :

$$h = b_0 + b_1 d + b_2 d^2$$

On notera, à propos de cette régression, les points suivants :

- [1] La méthode d'ajustement permettant d'estimer b_0 , b_1 , b_2 est expliquée en annexe A22, avec un exemple à l'appui.
- [2] S'il y a plusieurs placettes sur un même peuplement, on doit ajuster la régression hauteur/diamètre sur l'ensemble des arbres-échantillons.
- [3] On ne doit jamais regrouper des échantillons provenant de peuplements de densités et d'âges différents pour ajuster la régression, sauf si la comparaison des régressions séparées (voir Annexe A28) montre qu'il n'y a pas de différence significative.
- [4] On ne doit utiliser la régression ajustée pour prédire les hauteurs que si trois conditions sont satisfaites :
 - (a) la valeur de la statistique de Fisher (F-observé) de la régression est significative au seuil de 5 %
 - (b) le coefficient b_1 est positif
 - (c) le coefficient b_2 est négatif.

- 5 Une fois la régression calculée et testée, on peut estimer la hauteur moyenne du peuplement H par la valeur de h fournie par la régression pour $d = D_g$.
- 6 On peut de même estimer la hauteur dominante H_0 par la valeur de h fournie par la régression pour $d = D_0$ (diamètre moyen des 100 plus gros arbres à l'hectare).

Revenons maintenant au point [4] : si les trois conditions énumérées ne sont pas satisfaites, il vaut mieux admettre que la hauteur h de chaque arbre, utilisée pour l'estimation du volume, est égale à la hauteur moyenne H du peuplement. Le cas se présente dans l'étude de nombreuses essences tropicales, parce que les variations dans la hauteur des arbres ne sont pas nettement corrélées aux différences de diamètre, au point que les erreurs de mesure sont supérieures à l'influence du diamètre sur la hauteur. On trouvera plutôt les régressions significatives, traduisant une relation hauteur/diamètre bien marquée, chez les essences d'ombre avec d'assez fortes densités.

Dans les peuplements inéquiennes, la régression sera toujours significative et on pourra l'utiliser pour estimer la hauteur de chaque arbre. Mais il faut bien veiller à ne pas regrouper des données issues de peuplements différents pour en tirer une régression unique sans avoir vérifié leur homogénéité au moyen des tests statistiques appropriés.

Si l'on ajuste la régression hauteur/diamètre à une série de peuplements équiennes d'âges différents, elle sera également significative. Mais il faut bien noter qu'il s'agit là d'un modèle différent de celui que l'on ajuste à l'intérieur d'un peuplement équienne : il ne peut servir à estimer la hauteur de chaque arbre au sein d'un même peuplement équienne, mais on peut l'utiliser pour prédire la hauteur moyenne H en fonction du diamètre moyen D_g , à condition de rester dans une même zone de densité. Dans un peuplement équienne, la régression reflète des différences de vigueur ; entre classes d'âge différentes, elle exprime une croissance en fonction du temps.

CHAPITRE 4

SYSTEMES DE STOCKAGE DES DONNEES

4 SYSTEMES DE STOCKAGE DES DONNEES

41 AVANTAGES DES SYSTEMES DE STOCKAGE DES DONNEES SUR ORDINATEUR

Un système de stockage des données sur ordinateur permet d'enregistrer et de conserver sur bandes ou disques magnétiques toutes les données des placettes temporaires ou permanentes et des placettes d'expérience. Il permet d'accéder rapidement à ces données pour en tirer des résumés, les analyser, les soumettre à des procédures de détection automatique des erreurs, les mettre à jour ou les corriger sans difficulté.

Un tel système de stockage ne permet cependant pas de se passer d'un système de classement traditionnel où l'on conservera les fiches remplies sur le terrain, ainsi que les divers protocoles de mesure, notes, graphiques et autres documents afférents à chaque expérience ou placette permanente.

Jusqu'à ces derniers temps, le manque de personnel qualifié et l'importance du prix d'achat d'un ordinateur restreignaient considérablement l'utilisation de tels systèmes. Mais, d'une part, les prix ont considérablement baissés, jusqu'à être de l'ordre de ceux de véhicules à moteur, avec l'avènement des micro-ordinateurs ; d'autre part, ceux-ci sont généralement programmables en BASIC, langage conçu pour être appris facilement ; quiconque est doté d'un esprit normalement docile peut l'apprendre seul à condition d'avoir accès à un micro-ordinateur. Certains organismes forestiers ⁽¹⁾ offrent en outre aux ingénieurs et techniciens des stages de formation au traitement des données.

L'absence d'un système de stockage des données sur ordinateur constitue un grave handicap pour la construction et la validation de modèles de production, ainsi que pour un aménagement et un contrôle efficace de la production : on ne peut alors construire que les modèles les plus simples, et le contrôle de leur validité par analyse des résidus est véritablement impossible en raison de la somme de travail que cela implique. On ne peut pas non plus étudier diverses stratégies de modélisation sur les mêmes données, ni parfaire l'ajustement des fonctions constituant le modèle au fur et à mesure de la collecte de nouvelles données. On a enfin tendance à ne pas recueillir suffisamment de données, de peur de ne pas pouvoir les absorber. En outre, la productivité d'un personnel qualifié est considérablement diminuée par l'utilisation d'un système manuel.

C'est pourquoi nous recommandons à tout organisme forestier d'adopter l'une des deux solutions suivantes :

- a - Disposer des prestations de service d'un ordinateur de grande capacité, avec un délai maximum de traitement de un jour. Ce délai doit être compté du moment où les travaux partent de l'organisme au moment où les résultats y reviennent, et il faut prendre en considération les périodes du mois où l'ordinateur principal peut se trouver totalement indisponible (par exemple, parce qu'un droit d'utilisation prioritaire est accordé à

⁽¹⁾ par exemple le Commonwealth Forestry Institute d'Oxford, en Grande-Bretagne.

d'autres usagers), sans oublier les difficultés de transport ...etc,

- [b]- Acheter un système micro-ordinateur comportant :
- une mémoire de 32 à 64 k octets,
 - un système de deux disquettes ou disques rigides,
 - une imprimante,
 - un télétype pouvant être relié à l'ordinateur, avec lecteur-perforateur de ruban, pour la préparation des données "on line" ou "off line",
 - un interpréteur BASIC ou un compilateur FORTRAN.

Actuellement (été 1979), on peut acquérir une telle configuration pour environ 10 000 dollars U.S. en Europe ou en Amérique du Nord. Avec la solution [b] l'organisme aura un ordinateur sur place, à son entière disposition, ce qui permettra une meilleure productivité du personnel et une formation plus rapide qu'avec la solution [a].

42 CONTROLE DES DONNEES

Les erreurs qui affectent les données stockées sur bande ou disque magnétique peuvent provenir :

- d'une erreur de mesure sur le terrain,
- d'une erreur à la perforation ou à l'entrée des données
- d'une erreur de programmation.

On doit supposer que ce dernier type d'erreur est éliminé par une vérification méticuleuse de tous les programmes utilisés pour enregistrer, corriger, mettre à jour, imprimer ou extraire des données ; cette vérification est de la responsabilité directe et entière du programmeur, qui doit livrer des programmes sans aucune erreur.

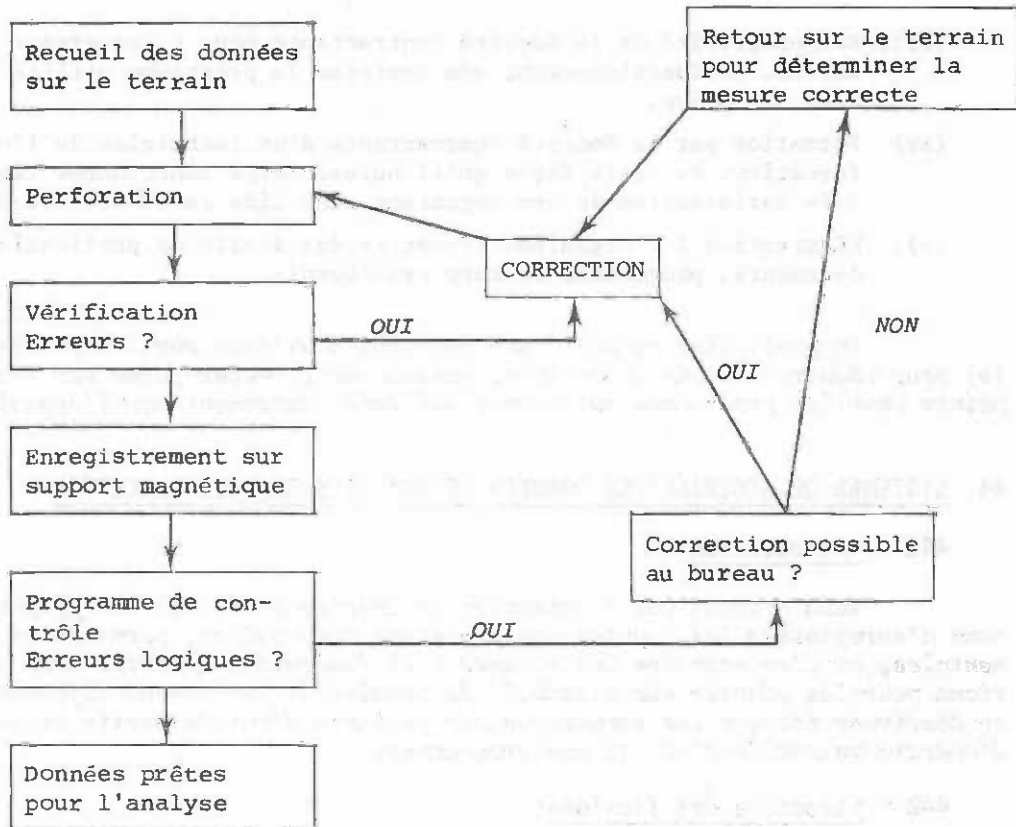
Les erreurs à la perforation ou à l'entrée des données sont presque entièrement éliminées si on adopte la procédure de vérification, obligatoire quand on traite une grande masse de données. Cette procédure consiste à faire entrer deux fois les données par le même opérateur, ou une fois par deux opérateurs différents : l'identité des deux jeux de données est automatiquement vérifiée, et l'opérateur apporte immédiatement la correction appropriée à toute incohérence révélée par cette vérification. Pour la plupart des ateliers de préparation des données, la vérification fait normalement partie du service offert, mais ce point doit toujours être explicitement précisé au moment de la remise des données à perforer.

Il n'est pas possible d'éliminer toutes les erreurs de mesure sur le terrain, mais on peut les réduire au minimum :

- [a] en veillant à la formation technique et psychologique des exécutants et en leur fournissant de bons instruments,
- [b] en soumettant les données, après leur enregistrement sur support magnétique, à des programmes de contrôle. Le principe de ce contrôle est la recherche de toutes les incohérences logiques du genre : mesures manifestement trop grandes ou trop petites, accroissements en hauteur ou en diamètre négatifs, arbres sans numéro d'identification, ou qui réapparaissent après avoir été récoltés, ou qui changent d'essence...etc.

Les incohérences ainsi révélées doivent être examinées pour déterminer l'origine de l'erreur. On corrige ensuite ces erreurs sur le support magnétique à l'aide d'un programme de mise en forme des données.

L'organigramme ci-dessous décrit l'ensemble du processus de contrôle des données :



43 LES CONTRATS DE FOURNITURE DE PROGRAMMES D'ORDINATEUR

Les contrats de fourniture de programmes d'ordinateur doivent comporter les clauses suivantes :

- (i) Fourniture par la Société contractante du listage des programmes-sources (en BASIC, FORTRAN ou autre langage à préciser), pour tous les programmes fournis.
- (ii) Fourniture d'une notice complète sur tous les programmes fournis, comportant :
 - a - un dictionnaire donnant la définition de tous les noms des variables utilisées dans le programme,
 - b - les organigrammes décrivant la séquence des opérations réalisées par le programme (et les sous-programmes), et la nature des trans-

ferts des données de et vers les supports externes et les fichiers magnétiques,

a - la définition précise de la structure des enregistrements et de l'usage de tous les fichiers magnétiques et de tous les fichiers d'entrée-sortie,

d - l'explication de la théorie sur laquelle s'appuie l'algorithme du programme, accompagnée de références bibliographiques ou autres,

(iii) Responsabilité de la Société contractante pour toute erreur de programmation, ou fonctionnement non conforme du programme utilisé en respectant la notice.

(iv) Formation par la Société contractante d'un technicien de l'organisme forestier, de telle façon qu'il puisse faire fonctionner les programmes à la satisfaction de son organisme sans aide de la Société contractante.

(v) Réservation à l'organisme forestier des droits de publication de tous les documents, programmes et rapports fournis.

On peut faire preuve d'une certaine souplesse sur les points (iii) et (v) pour réduire le coût du contrat, mais on doit rester ferme sur les autres points pour les programmes qui seront utilisés couramment par l'organisme forestier.

44 SYSTEMES DE STOCKAGE DES DONNEES RECUEILLIES SUR LES PLACETTES

441 Introduction

Nous n'avons pas l'intention de décrire en détail les programmes permettant d'enregistrer les données des placettes temporaires, permanentes ou expérimentales, et d'en extraire des résumés ; il faudrait y apporter trop de modifications pour les adapter aux possibilités locales de traitement informatique. Nous ne décrirons ici que les structures des fichiers d'entrée-sortie et les types d'opérations réalisables par ces programmes.

442 Structure des fichiers

En informatique, un fichier est un ensemble d'informations lisibles par une machine enregistrées sur un support magnétique (bande ou disque), sur cartes perforées ou sur ruban perforé. Les fichiers d'entrée fournissent des données au programme qui se déroule ; ils peuvent d'ailleurs avoir été créés par un autre programme, dont ils constituaient les fichiers de sortie.

Pour nous, la structure d'un fichier définira l'organisation des diverses informations à regrouper pour avoir les données concernant une placette.

Dans le cas des placettes permanentes, on peut définir ainsi l'information de base relative à une placette :

autant de fois
qu'il y a de
placettes per-
manentes.

{ - installation de la placette p
- 1ère mesure de la placette p
- 2ème mesure de la placette p
..... etc

{ - n^{ème} mesure de la placette p
- fin de la placette p

Dans le cas des placettes temporaires, la structure est beaucoup plus simple :

- mesure de la placette 1
- mesure de la placette 2
- etc
- mesure de la placette m

Ces structures s'appliquent aux enregistrements de la base de données permanente, qui peut être sur disque ou bande magnétique. Pour les placettes permanentes, il y a deux types d'enregistrements :

- (i) installation d'une placette
- (ii) mesure d'une placette.

Le premier crée une nouvelle placette dans la base de données : le programme doit donc vérifier que le numéro d'identification n'a pas déjà été attribué à une autre placette. Le second implique que la placette existe déjà : le programme doit le vérifier et stocker les données de la mesure à la bonne place. Lors de l'installation et de la première mesure d'une placette, on a affaire successivement à ces deux types d'enregistrement.

Rien n'empêche de traiter de la même façon les placettes expérimentales. On tâchera en outre de placer côte à côte sur la bande magnétique les placettes correspondant à des traitements voisins, par exemple :

Bloc 1 { - traitement 1
.....
- traitement n

Bloc 2 { - traitement 1
.....
- traitement n

..... etc

Bloc m { - traitement 1
.....
- traitement n

Il suffira alors, pour récupérer toutes les données de l'expérience, de lire $n \times m$ placettes à partir du début du fichier (première placette de l'expérience) sur la bande ou le disque.

443 Programmes de contrôle et de mise en forme des données

Les contrôles à faire sur les données des placettes permanentes sont les suivants :

- 1 que la placette existe déjà sur la bande (quand il ne s'agit pas d'ajouter une nouvelle placette à la base de données), ou qu'elle n'existe pas encore (quand il s'agit d'ajouter une nouvelle placette),
- 2 que les dates des mesures successives forment bien une suite cohérente,
- 3 que les arbres enlevés en éclaircie, qui ont donc disparu lors de mesures précédentes, ne réapparaissent pas dans les mesures ultérieures,
- 4 que les arbres (dans les peuplements hétérogènes), ou les placettes (dans les plantations) ne changent pas d'essence,
- 5 que les accroissements en diamètre sont positifs et inférieurs à une limite raisonnable,
- 6 que les valeurs de variables telles que hauteur, diamètre, ...etc soient d'un ordre de grandeur vraisemblable, et que les codes employés font bien partie de la liste des codes prévus.

Sur les placettes temporaires, on ne peut procéder qu'aux contrôles énumérés en 6.

Toutes les données sont ajoutées à la base de données, même si elles comportent des erreurs : on utilisera ensuite un programme de mise en forme pour les manipuler et les corriger dans la base. Dès que la masse des données devient importante, il est plus pratique de procéder ainsi que d'essayer de corriger les cartes perforées.

444 Résumés de placette

Le rôle principal du programme "résumés de placettes", qui peut d'ailleurs être associé au programme de contrôle, est de fournir, pour chaque placette, la liste des données qui serviront aux analyses ultérieures ; cette liste est un condensé de l'information recueillie sur la placette et son contenu dépend du type de forêt étudié et du type de modèle envisagé. Par exemple, s'il s'agit de plantations, le résumé de placette pourra contenir l'âge du peuplement, les hauteurs dominante et moyenne, la surface terrière, le nombre de tiges, le diamètre de l'arbre de surface terrière moyenne, et le volume. Dans les forêts hétérogènes, il contiendra aussi généralement des tableaux de peuplement donnant la valeur de certaines caractéristiques par catégories d'essences et de grosseur.

Le programme "résumés de placette" doit fournir ses résultats sous deux formes : une forme imprimée avec des titres pour en faciliter la lecture, et une forme enregistrée, sans textes, qui pourra directement servir de fichier d'entrée aux programmes d'analyse des données. Cette forme enregistrée peut l'être sur bande ou disque magnétique, sur cartes perforées ou ruban perforé, selon les moyens dont on dispose. Si l'on adopte une sortie sur support magnétique, ce sera sous la forme d'enregistrements "formatés" où chaque caractère (chiffre ou lettre) est codé séparément plutôt que sous la forme d'enregistrement "sans format" où chaque nombre (et non plus chaque chiffre) est codé selon la représentation interne des nombres propres à l'ordinateur.

Pour les forêts hétérogènes en particulier, mais aussi pour les peuplements équiennes monospécifiques, il est utile de pouvoir n'imprimer et n'enregistrer que certains résultats du résumé.

445 Autres programmes utilitaires

La gestion d'une base de données de placettes d'échantillonnage impliquera l'utilisation de deux autres programmes :

- un programme de tri capable d'ordonner, sur une bande magnétique, les placettes selon n'importe quelle "clé" de tri et permettant par exemple de les regrouper par forêt, série, essence, parcelle, ...etc. En effet, les placettes sont généralement enregistrées dans la base de données au fur et à mesure de leur installation, alors que l'on préfère obtenir les résumés dans un ordre logiquement défini.
- un programme de codage caractère par caractère et d'archivage, permettant de transférer la base de données principale implantée sur disque ou bande magnétique sous une forme convenant à l'archivage (voir ci-dessous) ou à la transmission à un autre centre de traitement ; ce programme réalisera donc l'enregistrement formaté, sur bande magnétique, de la totalité de la base de données.

446 Protection de la base de données

Une base de données enregistrée sur bande ou disque magnétique peut facilement être détruite par un accident matériel, ou sur une erreur du programme ou du système de l'ordinateur. C'est pourquoi il faut absolument, après chaque enregistrement d'un nombre important de données nouvelles, copier la base de données dans sa totalité sur un autre disque ou une bande, pour en avoir un exemplaire de rechange complet. On peut organiser une rotation des copies ainsi obtenues de façon à disposer à tout moment d'une version à jour servant au travail courant, d'une version archivée à jour et des deux ou trois versions archivées les plus récentes. Il faut éviter de les stocker toutes au même endroit.

45 TRANSFERT DE DONNEES ENTRE SYSTEMES DE TRAITEMENT

Il arrive fréquemment que l'on soit obligé de transférer des données d'un système de traitement à un autre, par exemple lors d'une recherche en coopération ou pour permettre à des chercheurs d'aller traiter leurs données ailleurs que dans l'organisme dont ils dépendent.

Si les données sont en quantité importante, le mieux est de les transférer sur des bandes magnétiques conformes aux normes industrielles, à 7 ou 9 pistes. Lors de l'enregistrement de la bande, on doit en préciser les spécifications :

- . nombre de pistes (7 ou 9)
- . densité (généralement 800 ou 1600 bits par inch)
- . parité (pair ou impair)
- . espace inter-blocs en mm (ou, à défaut, référence exacte du lecteur enregistreur de bandes utilisé).

Les bandes destinées au transfert de données entre ordinateurs doivent toujours porter des enregistrements formatés, de longueur fixe, si possible inférieure à 120 caractères pour en faciliter la lecture. Si on le peut, on doit préciser le type de code-caractère utilisé (EBCDIC, ASCII, DCB, ...), mais ce n'est pas fondamental, car la transcription de l'un dans l'autre est très facile. Il est bon

d'envoyer avec la bande un listage des premières et dernières centaines de lignes : ceci permet de vérifier, au moment de la lecture, qu'aucun enregistrement ou partie d'enregistrement n'a été perdu.

Les "labels" de tête et les "tape-marks" sont plus un inconvénient qu'une aide quand on lit une bande enregistrée ailleurs : il vaut mieux enregistrer toute l'information comme un seul fichier, sans "label" de bande au début, et la clore par une double "tape-mark".

On inscrira sur une étiquette collée au boîtier de la bande les indications concernant les spécifications de la bande (nombre de pistes, densité, parité, code-caractère), son contenu et l'adresse de l'organisme expéditeur. Rien ne s'oppose à un envoi par la poste, sinon que les détecteurs de métal à haute fréquence peuvent effacer tout ou partie de l'information : il faut donc indiquer clairement de quoi il s'agit sur le paquet, et faire en sorte que l'on puisse l'ouvrir facilement pour vérification.

Le tableau ci-dessous résume les résultats obtenus pour les différents peuplements homogènes étudiés. Les données relatives à l'accroissement et à la production sont présentées en fonction de la durée de culture et de la densité de peuplement. Les résultats indiquent que l'accroissement et la production augmentent avec la durée de culture et la densité de peuplement, mais que ces augmentations tendent à se stabiliser après une certaine durée de culture.

CHAPITRE 5

**ANALYSE DES DONNÉES
D'ACCROISSEMENT ET DE PRODUCTION
(PEUPEMENTS HOMOGENES)**

5 ANALYSE DES DONNEES D'ACCROISSEMENT ET DE PRODUCTION (PEUPELEMENTS HOMOGÈNES)

Dans un peuplement homogène, les arbres du peuplement principal ont tous le même âge et cet âge est connu ; ils appartiennent en outre à une même essence, ou à quelques essences écologiquement proches. Les peuplements homogènes au sens strict sont régénérés par coupe rase, éventuellement après une ou plusieurs éclaircies intermédiaires ; la régénération peut être obtenue par plantation, par semis (naturel ou artificiel) ou par rejets.

Dans ce type de peuplements, on sait bien quels sont les paramètres à prendre en compte dans la prévision de l'accroissement et de la production, et on a le choix entre un grand nombre de modèles possibles. Le facteur limitant dans l'efficacité d'un modèle donné est avant tout la possibilité de disposer, pour l'essence et la région étudiées, d'une masse de données couvrant une gamme étendue de stations, d'âges et de densités du peuplement.

L'accroissement et la production d'un peuplement peuvent être modélisés à trois niveaux différents ; l'ensemble du peuplement, la catégorie de grosseur, l'arbre pris individuellement. Dans les peuplements homogènes, les modèles de peuplement, plus faciles à construire et à utiliser, sont généralement satisfaisants : c'est donc d'eux que nous traiterons le plus longuement dans ce chapitre, les deux autres approches étant seulement présentées brièvement au paragraphe 57.

Même en se restreignant à ce niveau de modélisation, le choix des fonctions à incorporer dans le modèle reste d'une grande variété. Les paragraphes qui suivent exposent quelques possibilités et tentent de définir les situations auxquelles chacune s'applique le mieux ; il est impossible de les présenter toutes sans donner trop d'ampleur à ce manuel. Dans l'ensemble, les techniques que nous avons retenues l'ont été pour leur simplicité, leur précision et leur souplesse.

Nous avons essayé, pour chaque aspect particulier de la modélisation (par exemple : construction d'un jeu de courbes surface terrière/hauteur/densité du peuplement), d'exposer à la fois des méthodes graphiques et des méthodes statistiques plus complexes.

51 CLASSEMENT DE LA STATION (DETERMINATION DE LA FERTILITE) ⁽¹⁾

La précision que l'on peut obtenir dans l'application d'un modèle de production pour peuplements homogènes dépend en partie de la précision avec laquelle on peut classer la station. Celle-ci dépend à son tour de la précision sur l'âge du peuplement, généralement connu à partir des documents de gestion, et sur la hauteur des arbres dominants, mesurable à l'aide de dendromètres ou d'instruments analogues.

⁽¹⁾ en anglais = site classification

511 Utilisation de la hauteur dominante comme indicateur de fertilité

La hauteur atteinte par un peuplement homogène à un âge donné sur une station donnée est un bon indicateur de la productivité de ce type de peuplement sur ce type de station. C'est pourquoi la construction des courbes hauteur/âge correspondant aux différentes classes de fertilité constitue la première étape de la construction d'un modèle d'accroissement et de production.

La hauteur moyenne d'un peuplement étant généralement sensible, non seulement à l'âge et à la fertilité, mais aussi à la densité, on préfère définir la hauteur d'un peuplement par sa hauteur dominante, presque totalement insensible aux variations de densité.

La hauteur dominante peut être définie de plusieurs façons, mais la plus courante est la suivante : hauteur moyenne des 100 plus gros arbres à l'hectare.

Il arrive parfois, sous les tropiques, que la hauteur dominante d'un peuplement homogène ne soit pas un bon indicateur de la fertilité : cela arrive dans de jeunes peuplements à croissance très rapide, et avec certaines essences comme le *Pinus caribaea* qui font preuve d'une grande variabilité dans leur croissance en hauteur. On peut détecter ce phénomène en rangeant par hauteur dominante décroissante, à l'intérieur de chaque classe d'âge, les différentes mesures de placettes permanentes dont on dispose : si les rangs des placettes, au fil des mesures successives, sont peu corrélés entre eux, on ne peut pas construire sérieusement un faisceau de courbes de fertilité à partir de ces données.

Le problème vient simplement de la grande variabilité, sur une station donnée, de la croissance en hauteur des arbres de ces peuplements. On peut le résoudre en partie en redéfinissant la hauteur dominante sur un plus grand nombre d'arbres à l'hectare, par exemple 200 ou 400, et donc en mesurant la hauteur d'un plus grand nombre d'arbres-échantillons sur chaque placette. Une autre possibilité consiste à corréliser la productivité observée à certains facteurs du milieu comme la pente, l'altitude, le type de sol ou tout autre facteur se révélant significatif, la détermination de la fertilité se faisant alors à partir des valeurs observées de ces facteurs.

512 Construction des faisceaux des courbes de fertilité

La relation hauteur/âge/indice de fertilité constitue un élément fondamental de la prévision de l'accroissement des peuplements homogènes. On l'exprime habituellement sous la forme d'un faisceau de courbes de fertilité ⁽¹⁾ pour une essence donnée dans une région donnée.

On peut les construire par des méthodes graphiques ou statistiques (régression).

(1) en anglais = site index curves

512.1 Méthodes graphiques

Les étapes de la construction d'un faisceau de courbes de fertilité par méthode graphique sont les suivantes :

- 1 Porter sur un graphique hauteur-âge les données recueillies sur des peuplements de l'essence étudiée ; c'est la hauteur dominante qui est utilisée, et non la hauteur moyenne beaucoup plus dépendante de la densité. Une placette temporaire y sera représentée par un point, une placette permanente par une série de points reliés par des segments de droite. La figure 5.1 illustre cette première étape, avec des données provenant de peuplements de *Pinus Patula* en Ouganda.
- 2 Tracer ensuite à la main, dans le nuage de points, trois courbes représentant au mieux l'évolution moyenne suggérée par :
 - (i) les placettes du bord inférieur du nuage
 - (ii) la "tendance moyenne" du nuage
 - (iii) les placettes du bord supérieur du nuage.

Chacune de ces courbes doit s'appuyer au maximum sur l'évolution des placettes permanentes correspondantes. La figure 5.2 illustre le tracé de ces trois courbes.

Quand toutes les données proviennent de placettes temporaires, on ne peut plus se guider sur l'évolution des placettes permanentes : il y a alors un risque de biais important si les placettes temporaires observées n'échantillonnent pas avec la même intensité toutes les classes d'âge et de fertilité.

- 3 On peut alors intercaler deux autres courbes, l'une entre la courbe supérieure et la courbe centrale, l'autre entre la courbe centrale et la courbe inférieure. On reporte ensuite l'ensemble du faisceau sur un autre graphique pour obtenir la figure 5.3.
- 4 Le faisceau de courbes peut être traduit en équation grâce aux méthodes décrites en annexe A1. Les courbes elles-mêmes peuvent être numérotées en séquence, chacune représentant alors une classe de fertilité ⁽¹⁾ (ou de productivité). Ainsi, sur la figure 5.3, on a numéroté les classes de fertilité de I (la plus productive) à V (la moins productive).

La simplicité de cette méthode de construction graphique des courbes de fertilité est évidente. Mais elle présente trois inconvénients majeurs :

⁽¹⁾ en anglais = site class

Figure 5.1

Données provenant de placettes d'échantillonnage permanentes et temporaires, portées sur un graphique hauteur-âge en vue de la construction de courbes de fertilité

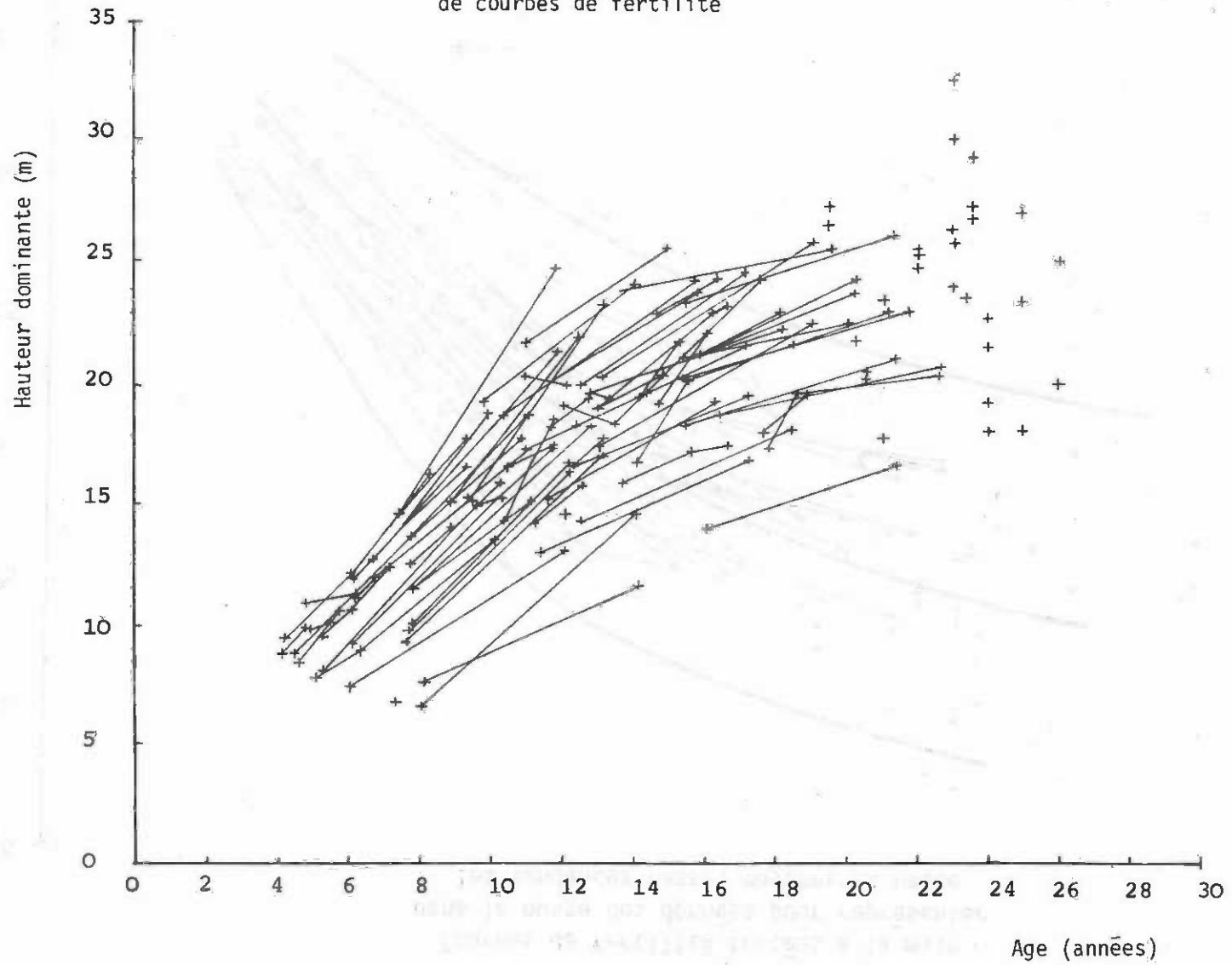


Figure 5.2

Courbes de fertilité tracées à la main dans le nuage des données pour représenter les tendances basse, moyenne et haute

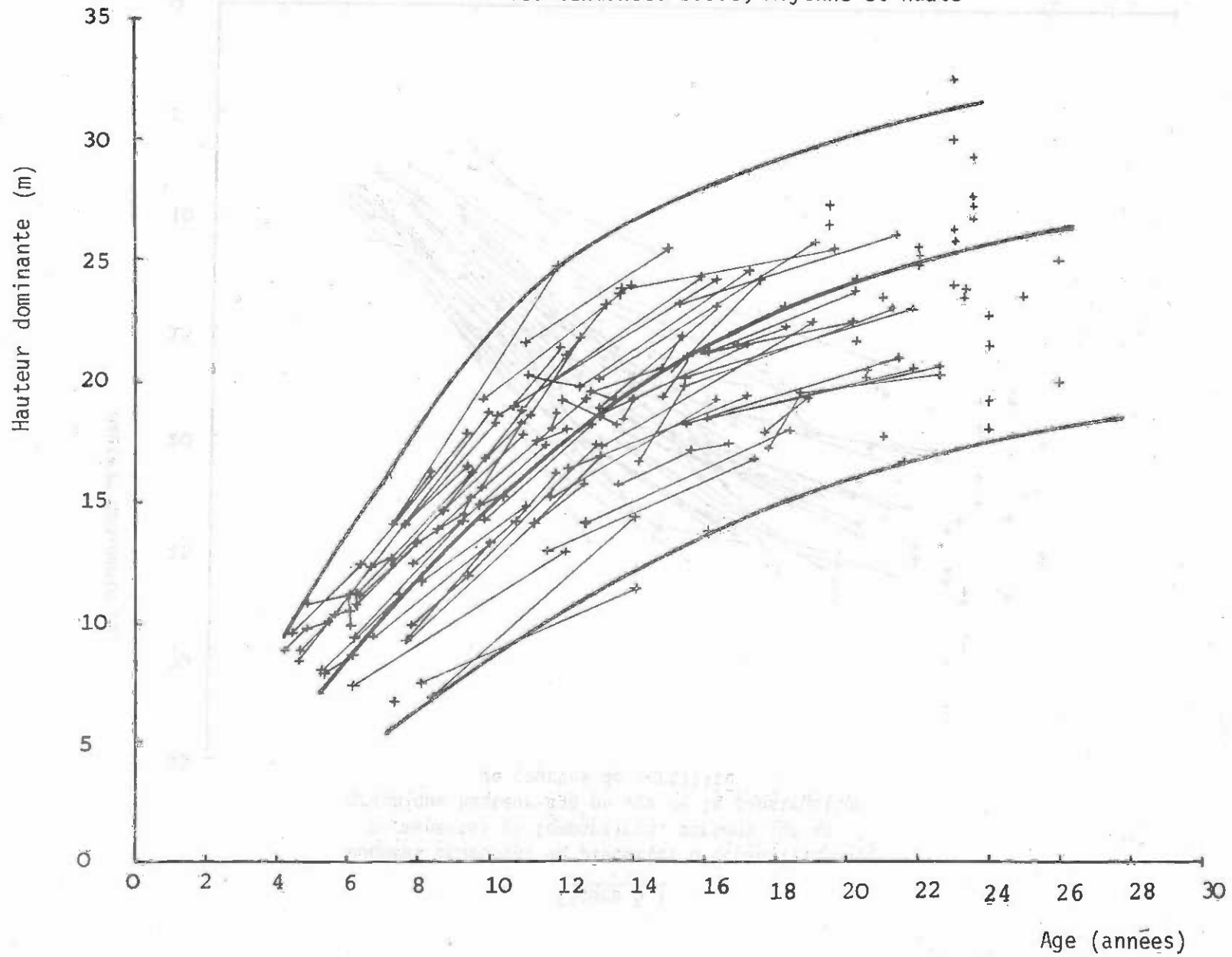
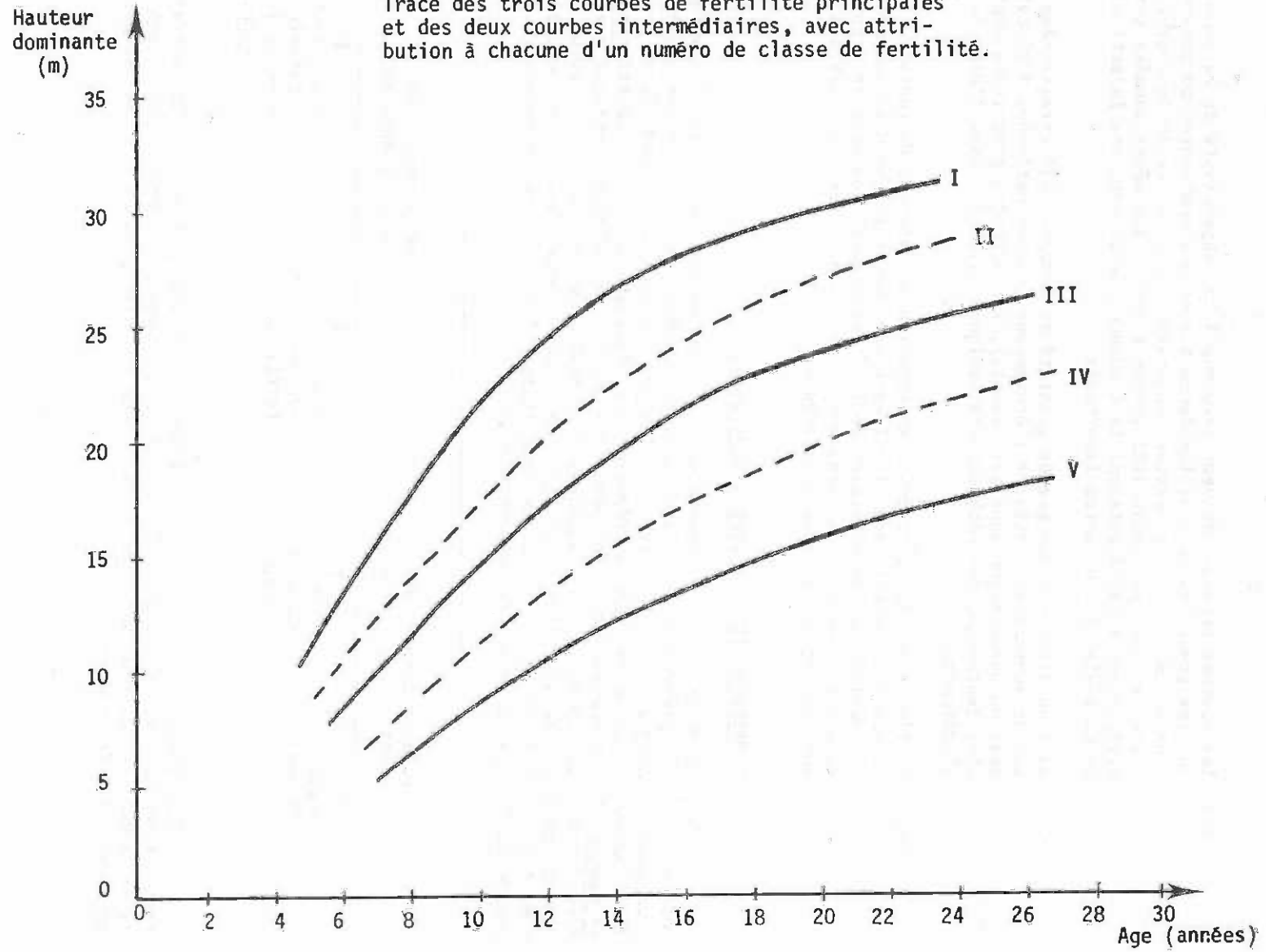


Figure 5.3

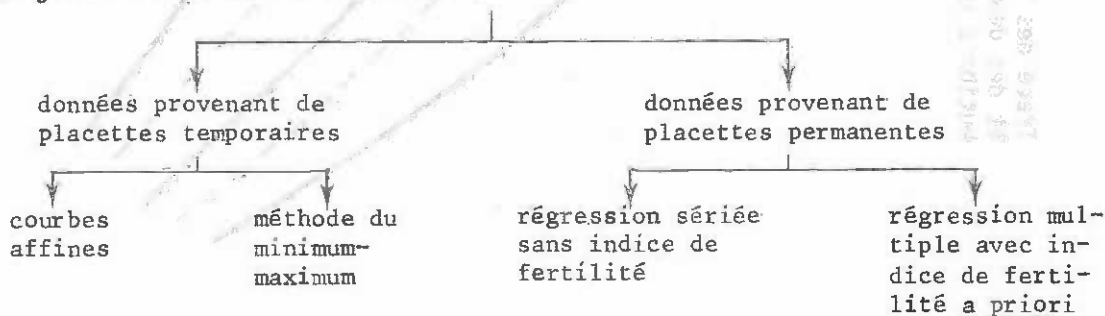
Tracé des trois courbes de fertilité principales et des deux courbes intermédiaires, avec attribution à chacune d'un numéro de classe de fertilité.



- (i) les courbes obtenues doivent beaucoup à la subjectivité de la personne qui les trace, surtout si les données sont peu nombreuses ou proviennent en majorité de placettes temporaires. Les faisceaux de courbes tracés par des personnes différentes à partir des mêmes données seront différents et représenteront la tendance réelle avec une incertitude et un biais plus ou moins importants.
- (ii) si l'on dispose d'une grande quantité de données déjà enregistrées sur un support accessible à l'ordinateur (cartes perforées à 80 colonnes, ou bande magnétique par exemple), les méthodes graphiques sont plus lentes que des méthodes statistiques utilisées avec l'aide de l'ordinateur.
- (iii) Si l'on a besoin de traduire en équation le faisceau de courbes obtenu (par exemple pour l'utiliser dans un programme d'actualisation d'inventaire ou de prévision de l'accroissement), on peut se trouver devant un travail aussi long que l'ajustement direct de courbes aux données par une méthode statistique.

512.2 Méthodes statistiques d'ajustement

On peut aussi utiliser, pour ajuster les courbes de fertilité, des méthodes mathématiques qui présentent des avantages considérables sur les méthodes graphiques quand on dispose d'un ordinateur et d'un grand nombre de données. Il ne faudrait cependant pas croire que la précision des résultats ainsi obtenus est nécessairement plus grande que celle des résultats obtenus par méthode graphique ; elle dépend pour beaucoup de la bonne adaptation du modèle de croissance en hauteur choisi, ainsi que de la validité des hypothèses statistiques faites lors de l'ajustement des paramètres de ce modèle. On peut classer les techniques mathématiques en quatre catégories de la manière suivante :



La régression sériée est, d'un point de vue statistique, la plus satisfaisante de ces quatre méthodes ; elle a aussi l'avantage de se prêter facilement au calcul manuel. C'est pourquoi nous l'étudierons en détail, en passant un peu plus rapidement sur les autres.

On peut utiliser les quatre méthodes avec un même modèle de croissance en hauteur : l'équation de SCHUMACHER ⁽¹⁾ :

$$H_0 = H_{\max} \times \exp(b/A^k) \quad (1)$$

où H_0 est la hauteur dominante,

H_{\max} un paramètre à ajuster qui représente la hauteur maximum que peut atteindre l'essence,

$\exp(\)$ la notation de la fonction exponentielle de constante $e = 2,71828$ (soit $\exp(x) = e^x = \text{"e puissance x"}$)

b et k des paramètres à ajuster,

A l'âge du peuplement.

La figure 5.4 montre la forme des courbes correspondant à cette équation. En prenant le logarithme à base e (logarithme népérien noté \ln ou \log_e) des deux membres de l'équation (1), on obtient :

$$\ln H_0 = \ln H_{\max} + b/A^k \quad (2)$$

En posant $a = \ln H_{\max}$, on peut ajuster a et b par régression linéaire à condition que la valeur de k soit fixée. Pour la plupart des essences, la valeur convenable de k se situe entre 0,2 et 2 : on peut l'estimer par des techniques décrites plus loin dans ce chapitre, ou par une méthode d'ajustement non linéaire présentée en annexe A4. Pour de nombreuses essences, une valeur de $k = 1$ choisie a priori conduira à des ajustements satisfaisants. Dans l'équation (2), le paramètre b doit toujours prendre une valeur négative, sinon il faut vérifier les calculs. La valeur du paramètre a se situe généralement entre 2 et 7 : si elle s'éloigne de cette fourchette, il faudra également vérifier les calculs.

512.21 La méthode des courbes affines

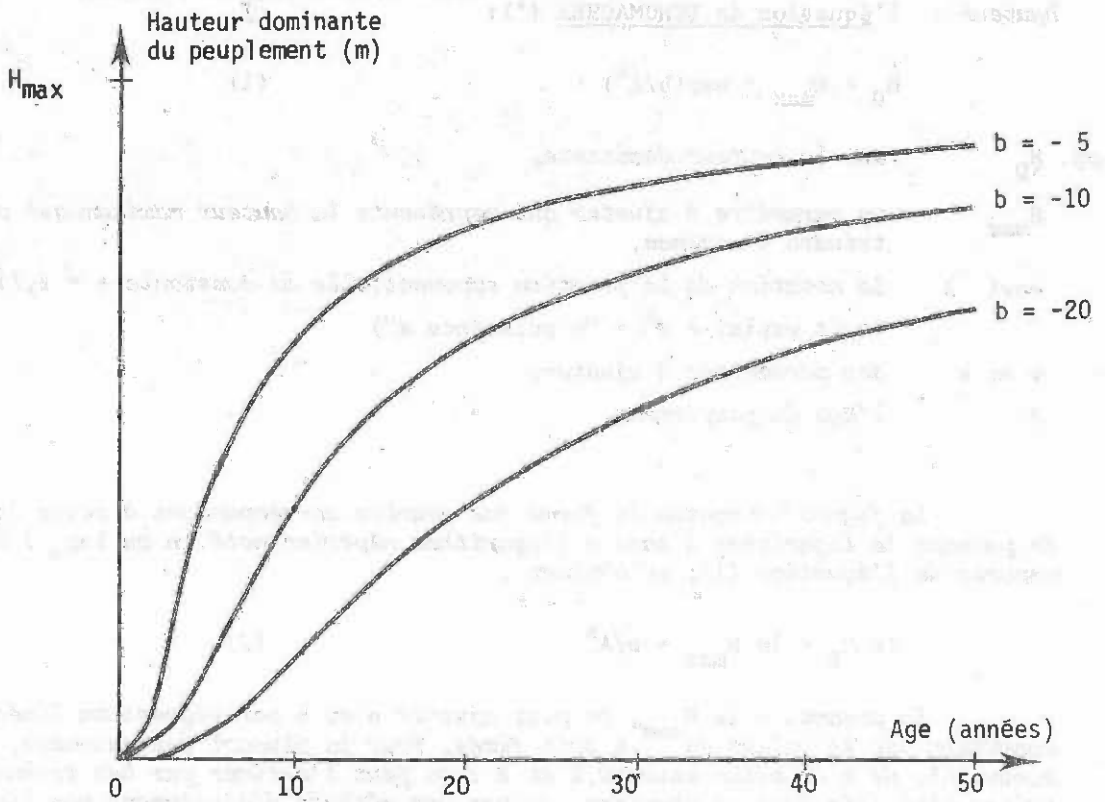
Dans cette première méthode, on ajuste par régression linéaire l'équation (2) à l'ensemble des données des placettes temporaires ; la variable expliquée Y est $\ln H_0$, la variable explicative X est $1/A^k$ (si la valeur de k n'est pas connue, voir en annexe A4 comment la déterminer).

Si l'on peut supposer que toutes les stations sont également représentées à l'intérieur de chaque classe d'âge, on obtient ainsi la tendance moyenne de la croissance en hauteur. Mais si l'on sait, par exemple, que les peuplements âgés sont plutôt sur des stations pauvres et les jeunes sur des stations fertiles, on doit renoncer à employer cette méthode au profit de méthodes graphiques ou, si on dispose de données provenant de placettes permanentes, de la régression sériée.

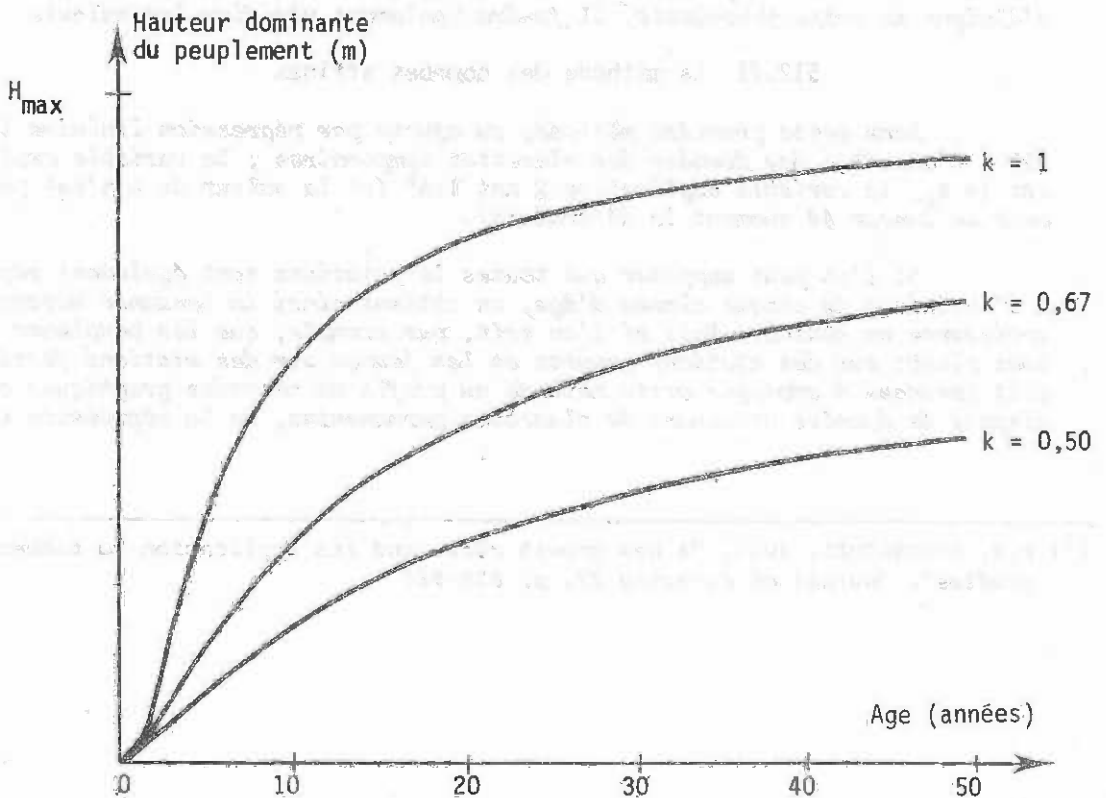
⁽¹⁾ F.X. SCHUMACHER, 1939, "A new growth curve and its application to timber yield studies", Journal of Forestry 37, p. 819-820

Figure 5.4.

a) Courbes correspondant à l'équation de Schumacher $H_0 = H_{\max} \times \exp(b/A^k)$ pour $k=1$ et différentes valeurs de b



b) Courbes correspondant à l'équation de Schumacher $H_0 = H_{\max} \times \exp(b/A^k)$ pour $b = -5$ et différentes valeurs de k



Une fois ajustée la courbe moyenne de croissance en hauteur, on peut tracer des courbes de même forme correspondant à différentes fertilités. On commence par définir un indice de fertilité ⁽¹⁾ H_r (hauteur de référence) qui est la hauteur dominante du peuplement à un âge de référence A_r . On dit ensuite que deux courbes ont la même forme si elles se déduisent l'une de l'autre par affinité sur l'axe des abscisses : si la première courbe a pour équation $y_1 = f(x)$, alors la seconde aura pour équation $y_2 = c \times f(x)$ où c est une constante. Pour l'équation de SCHUMACHER, on obtient donc des courbes affines en changeant simplement la valeur de H_{\max} , c'est-à-dire de $a = \ln H_{\max}$. On peut donc construire le faisceau de courbes de fertilité en retenant les valeurs de b et k obtenues pour la courbe moyenne et en prenant les valeurs de a correspondant à différentes valeurs de l'indice de fertilité : pour une valeur H_r de l'indice de fertilité, la valeur correspondante de a sera telle que :

$$\ln H_r = \ln H_{\max} + b/A_r^k = a + b/A_r^k$$

soit $a = \ln H_r - b/A_r^k$ (3)

512.22 La méthode du minimum-maximum

Cette deuxième méthode est plus souple que la méthode des courbes affines en ce qui concerne la forme des courbes, mais elle nécessite plusieurs observations (au moins 3) dans chaque classe d'âge, ce qui interdit son emploi quand on ne dispose que d'un nombre limité de données. Les étapes en sont les suivantes :

- [1] A l'intérieur de chaque classe d'âge, noter le minimum, le maximum et la valeur moyenne des hauteurs dominantes H_0 des placettes mesurées.
- [2] Ajuster trois régressions du type de l'équation (2) ci-dessus respectivement aux trois séries ainsi obtenues (minimum, moyenne, maximum). On peut admettre que le paramètre non linéaire k est le même pour les trois séries, mais on peut aussi l'ajuster indépendamment.
- [3] On peut enfin, en utilisant les méthodes décrites dans l'annexe A1 sous le nom d'harmonisation, lisser en fonction de l'indice de fertilité H_r (hauteur dominante à un âge de référence A_r) les coefficients obtenus séparément pour chacune des trois courbes : on obtient ainsi une équation unique du faisceau où les courbes sont paramétrées par H_r .

Si l'on dispose de nombreuses données dans chaque classe d'âge, comme en fournit un inventaire forestier, on utilisera plutôt une variante un peu plus complexe de cette méthode.

(1) en anglais = site index

Dans chaque classe d'âge, on range les placettes par ordre de hauteur décroissante, et on attribue à chaque placette un niveau de fertilité q selon la formule :

$$q = (i - 1/2)/n$$

dans laquelle i est le rang de la placette et n le nombre de placettes dans la classe d'âge. Quand toutes les placettes se sont vues ainsi attribuer un niveau de fertilité, on les analyse par régression multiple comme dans la dernière des quatre méthodes mathématiques, décrite au § 512.24.

Il faut bien voir que les méthodes décrites ci-dessus pour traiter les données provenant de placettes temporaires fournissent des résultats d'une utilité toute relative et provisoire, puisqu'ils sont fondés sur l'hypothèse que tous les types de station ont une probabilité égale d'être représentés dans chaque classe d'âge. Comme c'est rarement réalisé, les courbes obtenues seront plus ou moins défectueuses.

La seule solution satisfaisante consiste à tirer de l'observation de placettes permanentes ou d'analyses de tiges des séries de données hauteur/âge et à les analyser par l'une des méthodes qui suivent.

512.23 Les méthodes de régression sériée

Elles sont de deux types. Il y a d'abord la régression multiple avec utilisation de variables conditionnelles, telle qu'elle est décrite dans l'exemple de l'annexe A29. A notre connaissance, cette méthode n'a pas été utilisée pour construire des courbes de fertilité, probablement parce que le nombre de variables explicatives dans la régression devient énorme dès que le nombre des placettes atteint une valeur raisonnable ; mais cette approche n'a rien d'irréalisable en soi, avec un programme spécial pour créer et manipuler les nombreuses variables binaires 0/1. La seconde méthode, décrite pour la première fois par BAILEY et CLUTTER ⁽¹⁾, utilise les estimateurs "de pente unique" ⁽²⁾ et "de constante unique" ⁽³⁾ fournis par l'analyse de covariance. Cette méthode convient bien à la construction de courbes de fertilité et sa simplicité permet d'effectuer manuellement le calcul des paramètres.

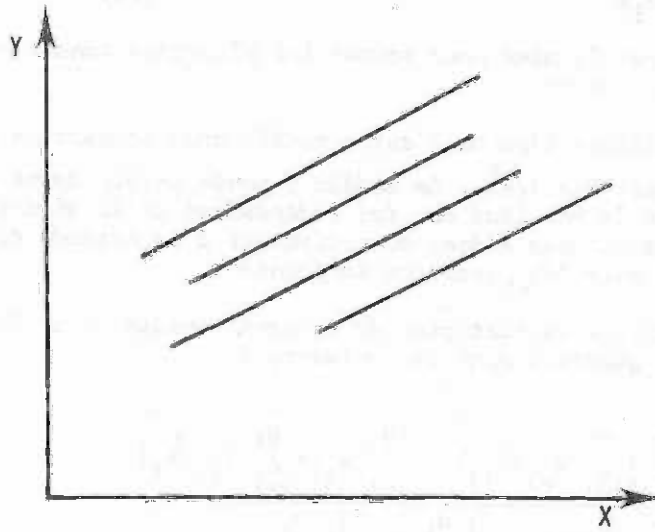
Le modèle de régression à pente unique, représenté sur la figure 5.5 (a), se traduit par l'équation :

$$Y = a_i + bX \quad (4)$$

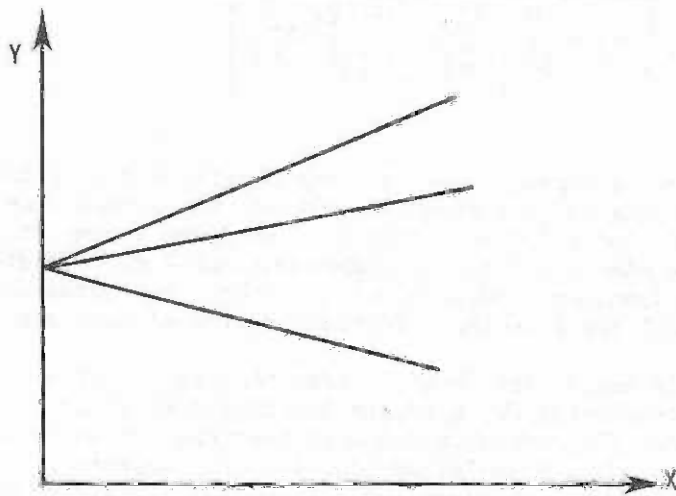
où la constante a_i varie d'une placette à l'autre, mais où la pente b est la même pour toutes les placettes.

Le modèle de régression à constante unique, représenté sur la figure 5.5 (b) se traduit par l'équation :

-
- (1) R.L. BAILEY et J.L. CLUTTER, 1974, "Base-Age invariant polymorphic site curves", Forest Science 20, pages 155-159
(2) en anglais = common slope estimator
(3) en anglais = common intercept estimator



a) Régressions à pente unique pour quatre placettes.



b) Régressions à constante unique pour trois placettes.

Figure 5.5

Régressions linéaires à pente unique ou à constante unique.

$$Y = a + b_i X \quad (5)$$

où la constante a est la même pour toutes les placettes tandis que la pente b_i change avec chaque placette.

On peut utiliser l'un ou l'autre modèle avec l'équation de SCHUMACHER en faisant $Y = \ln H_0$ et $X = 1/A^k$. Le modèle à pente unique donne des faisceaux de courbes affines mais le résultat est ici indépendant de la distribution des stations dans les différentes classes d'âge, contrairement à la méthode des courbes affines présentée ci-dessus pour les placettes temporaires.

Les estimateurs statistiques de la pente unique et de la constante unique dans les modèles en question sont les suivants :

$$\text{Pente unique } b = \frac{\sum_{i=1}^m \left[\sum_{j=1}^{n_i} (X_{ij} \times Y_{ij}) - \sum_{j=1}^{n_i} X_{ij} \times \sum_{j=1}^{n_i} Y_{ij} / n_i \right]}{\sum_{i=1}^m \left[\sum_{j=1}^{n_i} X_{ij}^2 - \left(\sum_{j=1}^{n_i} X_{ij} \right)^2 / n_i \right]} \quad (6)$$

$$\text{Constante unique } a = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} Y_{ij} - \sum_{i=1}^m \left[\sum_{j=1}^{n_i} X_{ij} \times \sum_{j=1}^{n_i} (X_{ij} \times Y_{ij}) / \sum_{j=1}^{n_i} X_{ij}^2 \right]}{\sum_{i=1}^m n_i - \sum_{i=1}^m \left[\left(\sum_{j=1}^{n_i} X_{ij} \right)^2 / \sum_{j=1}^{n_i} X_{ij}^2 \right]} \quad (7)$$

Ces formules pouvant paraître complexes (bien que le principe en soit simple), nous présentons ici des fiches de calcul, illustrées d'un exemple concret. Il s'agit de l'imprimé 5.1, fiches 1 et 2 ; la fiche 1 comporte les sommations sur les mesures d'une même placette, correspondant aux Σ sur l'indice j (variant de 1 à n_i) dans les formules ci-dessus, et la fiche 2 les sommations sur les placettes, correspondant aux Σ sur l'indice i (variant de 1 à m) dans ces formules.

L'exemple traite les données présentées sur la figure 5.6, qui proviennent de 6 placettes permanentes de *Cupressus lusitanica* au Kenya. Les données hauteur-âge correspondantes figurent dans les deux premières colonnes des tableaux de la fiche 1 (il en faut deux exemplaires pour les six placettes). Dans cet exemple, on a attribué arbitrairement au paramètre k la valeur 1 ; les valeurs des variables transformées X et Y , c'est-à-dire $1/A$ et $\ln H_0$, sont portées en colonnes 3 et 4, X^2 en colonne 5 et XY en colonne 6. Les calculs sont faits avec 4 chiffres significatifs au moins. La dernière ligne de chaque tableau contient le nombre de mesures de la placette et les sommes de chacune des colonnes 3 à 6 (sommes sur les mesures de la placette). Puis on passe à la fiche 2 de l'imprimé 5.1 pour la suite des calculs : on calcule pour chaque placette, à partir du nombre de mesures n et des sommes déjà calculées (ΣX , ΣY , ΣX^2 , ΣXY), les quantités indiquées en tête des colonnes de cette fiche. Ces quantités sont ensuite sommées sur les placettes, pour obtenir les sommes (1) à (6) en bas de colonne. On calcule enfin les estimateurs

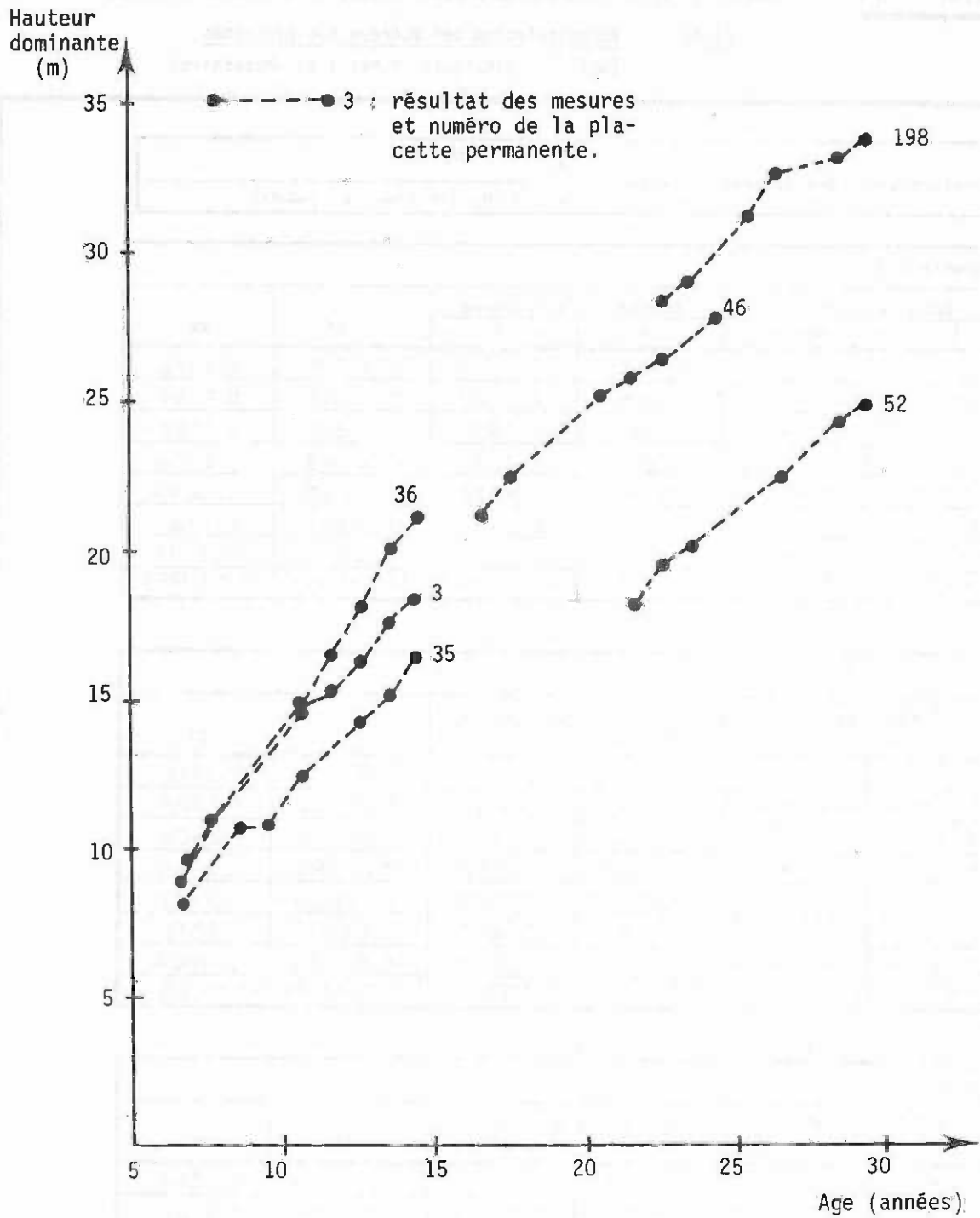


Figure 5.6

Données provenant de 6 placettes permanentes d'échantillonnage installées dans des peuplements de *Cupressus lusitanica* au Kenya, utilisées dans un exemple sur l'ajustement de courbes de fertilité par régression sériée.

Fiche 1 : Récapitulation des données des placettes
(utiliser plusieurs fiches 1 si nécessaire)

Transformations des données utilisées : $X = 1/AGE$
 $Y = LOG_e$ (Hauteur dominante)

Placette : 3

Données brutes		Données transformées		x ²	XY
A	H	X	Y		
6,5	9,6	0,1493	2,2618	0,02228	0,3376
7,6	11,0	0,1316	2,3979	0,01731	0,3155
10,6	14,8	0,09434	2,6946	0,00890	0,2542
11,6	15,2	0,08621	2,7213	0,00743	0,2346
12,6	16,3	0,07937	2,7912	0,00630	0,2215
13,5	17,5	0,07407	2,8622	0,00549	0,2120
14,4	18,3	0,06944	2,9069	0,00482	0,2019
Totaux	n = 7	$\Sigma X = 0,6843$	$\Sigma Y = 18,6358$	$\Sigma X^2 = 0,07253$	$\Sigma XY = 1,7773$

Placette : 35

Données brutes		Données transformées		x ²	XY
A	H	X	Y		
6,7	8,1	0,1493	2,0919	0,02228	0,3122
8,5	10,7	0,1176	2,3702	0,01384	0,2789
9,5	10,8	0,1053	2,3795	0,01108	0,2505
10,6	12,5	0,09434	2,5257	0,00890	0,2383
12,6	14,3	0,07937	2,6603	0,00630	0,2111
13,5	15,2	0,07407	2,7213	0,00549	0,2016
14,4	16,5	0,06944	2,8034	0,00482	0,1947
Totaux	n = 7	$\Sigma X = 0,6894$	$\Sigma Y = 17,5523$	$\Sigma X^2 = 0,07271$	$\Sigma XY = 1,6872$

Placette : 36

Données brutes		Données transformées		x ²	XY
A	H	X	Y		
6,7	8,9	0,1493	2,1861	0,02228	0,3263
7,6	11,1	0,1316	2,4069	0,01731	0,3167
10,6	14,6	0,09434	2,6810	0,00890	0,2529
11,6	16,5	0,08621	2,8033	0,00743	0,2417
12,6	18,1	0,07937	2,8959	0,00630	0,2298
13,5	20,0	0,07407	2,9957	0,00549	0,2219
14,4	21,0	0,06944	3,0445	0,00482	0,2114
Totaux	n = 7	$\Sigma X = 0,6843$	$\Sigma Y = 19,0135$	$\Sigma X^2 = 0,07253$	$\Sigma XY = 1,8007$

IMPRIME 5.1

MODELES DE REGRESSION SIMPLE A PENTE UNIQUE OU A CONSTANTE UNIQUE

Fiche 1 : Récapitulation des données des placettes
(utiliser plusieurs fiches 1 si nécessaire)

Transformations des données utilisées : $x = 1/AGE$
 $y = LOG_e$ (Hauteur dominante)

Placette : 46

Données brutes		Données transformées		x ²	xy
A	H	x	y		
16,7	21,1	0,05988	3,0493	0,00359	0,1826
17,6	22,4	0,05682	3,1091	0,00323	0,1767
20,6	25,1	0,04854	3,2229	0,00236	0,1564
21,6	25,7	0,04630	3,2465	0,00214	0,1503
22,6	26,3	0,04425	3,2696	0,001959	0,1447
24,4	27,7	0,04098	3,3214	0,001680	0,1361
Totaux	n = 6	$\Sigma x = 0,2968$	$\Sigma y = 19,2189$	$\Sigma x^2 = 0,01495$	$\Sigma xy = 0,9468$

Placette : 52

Données brutes		Données transformées		x ²	xy
A	H	x	y		
21,7	18,1	0,04608	2,8959	0,00212	0,1334
22,6	19,4	0,04425	2,9653	0,00196	0,1312
23,5	20,0	0,04255	2,9957	0,00181	0,1275
26,6	22,3	0,03759	3,1046	0,00141	0,1167
27,6	23,3	0,03623	3,1485	0,00131	0,1141
28,5	24,2	0,03509	3,1864	0,00123	0,1118
29,4	24,7	0,03401	3,2068	0,00116	0,1091
Totaux	n = 7	$\Sigma x = 0,2758$	$\Sigma y = 21,5031$	$\Sigma x^2 = 0,01100$	$\Sigma xy = 0,8438$

Placette : 198

Données brutes		Données transformées		x ²	xy
A	H	x	y		
22,7	28,3	0,04405	3,3429	0,00194	0,1473
23,5	28,9	0,04255	3,3638	0,00181	0,1431
25,6	31,1	0,03906	3,4372	0,00153	0,1343
26,5	32,5	0,03774	3,4812	0,00142	0,1314
28,6	33,0	0,03497	3,4965	0,00122	0,1223
29,4	33,6	0,03401	3,5145	0,00116	0,1195
Totaux	n = 6	$\Sigma x = 0,2324$	$\Sigma y = 20,6362$	$\Sigma x^2 = 0,00908$	$\Sigma xy = 0,7978$

de pente unique et de constante unique comme indiqué au bas de la fiche.

Dans l'exemple traité, on obtient finalement comme résultats :

- 9,222 pour l'estimateur de pente unique
- 3,583 pour l'estimateur de constante unique.

Si on trace sur un graphique le faisceau de courbes de fertilité correspondant à l'un ou l'autre de ces modèles, on s'apercevra dans les deux cas qu'il grimpe trop abruptement pour s'ajuster correctement aux données de la figure 5.6 : c'est que la valeur arbitrairement attribuée à k est beaucoup trop grande pour ces données.

BAILEY et CLUTTER montrent, dans l'article mentionné, comment on peut estimer directement le paramètre non linéaire k si l'on suppose qu'il a la même valeur pour toutes les placettes, en utilisant un modèle de régression où il figure en coefficient d'un terme linéaire. Il suffit pour cela de disposer de mesures successives (placettes permanentes ou analyses de tiges) permettant d'estimer des accroissements en hauteur. La méthode est la suivante :

- 1] Calculer les valeurs d'une nouvelle variable Y correspondant aux 2ème, 3ème, ... n_i^e observations d'une placette i , à partir de la formule ⁽¹⁾ :

$$Y_{ij} = \ln \left\{ \frac{\left[\frac{H_{ij} - H_{ij-1}}{A_{ij} - A_{ij-1}} \right]}{\left[\frac{H_{ij} + H_{ij-1}}{2} \right]} \right\}$$

Il n'y a pas de valeur de Y correspondant à la première mesure de hauteur.

- 2] Calculer de même les valeurs d'une nouvelle variable X à partir de la formule ⁽²⁾ :

$$X_{ij} = \ln \left(2 / (A_{ij} + A_{ij-1}) \right)$$

⁽¹⁾ Y_{ij} est le logarithme de l'accroissement moyen annuel relatif en hauteur entre les mesures $j-1$ et j sur la placette i . A la limite :

$$Y = \ln \left(\frac{1}{H} \frac{dH}{dA} \right)$$

soit avec l'équation de SCHUMACHER : $Y = \ln \left(\frac{-kb}{A(k+1)} \right) = \ln(-kb) + (k+1) \ln \left(\frac{1}{A} \right)$

⁽²⁾ X_{ij} est le logarithme de l'inverse de l'âge à mi-période entre les mesures $j-1$ et j sur la placette i . A la limite :

$$X = \ln \left(\frac{1}{A} \right)$$

La pente de la régression de la nouvelle variable Y sur cette nouvelle variable X fournit donc une estimation de $k+1$.

i	j	A _{ij}	a	b	c	X _{ij}	H _{ij}	d	e	f	g	h	Y _{ij}
Placette	Mesure	Age	Somme	Diffé- rence	$\frac{2}{a}$	ln c	Hauteur	Somme	Diffé- rence	$\frac{e}{b}$	$\frac{d}{2}$	$\frac{f}{g}$	ln h
4	1	16,7					21,1						
4	2	17,6	34,3	0,9	0,05831	-2,842	22,4	43,5	1,3	1,444	21,75	0,06639	-2,712
4	3	20,6	38,2	3,0	0,05236	-2,950	25,1	47,5	2,7	0,9000	23,75	0,03789	-3,273
4	4	21,6	42,2	1,0	0,04739	-3,049	25,7	50,8	0,6	0,6000	25,40	0,02362	-3,746
4	5	22,6	44,2	1,0	0,04525	-3,096	26,3	52,0	0,6	0,6000	26,00	0,02308	-3,769
4	6	24,4	47,0	1,8	0,04255	-3,157	27,7	54,0	1,4	0,7778	27,00	0,05185	-2,959

→ A reporter sur la fiche 1 de l'imprimé 5.1

- 3] Calculer l'estimateur de pente unique des régressions $Y = a_i + bX$ à l'aide de l'imprimé 5.1 ou à partir de l'équation (6). Remarque : si on utilise l'imprimé 5.1, on n'a besoin que des deux premières colonnes de la fiche 2.
- 4] Soustraire 1 de l'estimateur de pente unique ainsi obtenu pour avoir la valeur de k cherchée.

Pour calculer à la main les valeurs des nouvelles variables X et Y (étapes 1] et 2] ci-dessus), on peut s'aider de l'imprimé 5.2. Pour en illustrer l'emploi, nous avons rempli le premier cadre à partir des données de la quatrième placette de notre exemple (placette 46). Mais toutes ces formules (estimateurs de pente unique et de constante unique, calcul des nouvelles variables X et Y pour l'estimation de k) sont faciles à programmer sur un petit ordinateur, ou une calculatrice programmable ayant au moins 15 registres de mémoire et 200 pas de programme.

Quand on ajuste ainsi le paramètre k aux données de la figure 5.6 on obtient :

$$k = 0,25$$

et les estimateurs de pente unique et de constante unique deviennent :

$$b = - 6,638$$

$$a = 6,311$$

Voyons maintenant comment construire le faisceau de courbes de fertilité, à partir de ces résultats, en choisissant par exemple le modèle à constante unique.

L'équation de ces courbes est :

$$\ln H_0 = 6,311 + b_i / A^{0,25} \quad (8)$$

Le paramètre b_i dépend de l'indice de fertilité. Pour une valeur H_r de cet indice (hauteur atteinte à l'âge de référence A_r), on a :

$$\ln H_r = 6,311 + b_i / A_r^{0,25}$$

soit :

$$b_i = (\ln H_r - 6,311) \times A_r^{0,25}$$

et, si l'on veut tracer les courbes correspondant à des indices de fertilité de 16, 18, 20, 22, 24, 26 mètres à l'âge de référence de 20 ans :

H_r	b_i
16	- 7,483
18	- 7,234
20	- 7,011
22	- 6,809
24	- 6,625
26	- 6,456

En remplaçant b_i par sa valeur dans l'équation (8) pour chaque courbe de fertilité, on peut calculer les valeurs de H_0 correspondant à des valeurs données de A . La figure 5.7 représente les courbes ainsi obtenues à la même échelle que celle de la figure 5.6 qui présentait les données.

Si on veut maintenant déterminer l'indice de fertilité d'un peuplement, d'âge A et de hauteur dominante H_0 connus, on utilisera la formule :

$$\ln H_r = a + (\ln H_0 - a) \times (A/A_r)^k \quad (9)$$

Ainsi, avec le faisceau de courbes de fertilité ci-dessus, un peuplement dont la hauteur dominante est 14,5 m à 11,5 ans aura pour indice de fertilité :

$$\begin{aligned} \ln H_r &= 6,311 + (\ln 14,5 - 6,311) \times (11,5/20)^{0,25} \\ &= 3,144 \end{aligned}$$

soit $H_r = 23,2$ que l'on peut arrondir à 23.

Si l'on avait choisi le modèle à pente unique au lieu du modèle à constante unique, l'équation des courbes aurait été :

$$\ln H_0 = a_i + b/A^k \quad (10)$$

où a_i dépend de l'indice de fertilité ; pour une valeur H_r de cet indice :

$$a_i = \ln H_r - b/A_r^k \quad (11)$$

et l'indice de fertilité d'un peuplement d'âge A et de hauteur dominante H_0 serait donné par :

$$\ln H_r = \ln H_0 + b (1/A^k - 1/A_r^k) \quad (12)$$

512.24 Les méthodes de régression multiple

Ces méthodes sont utilisables pour l'ajustement d'un faisceau de courbes de fertilité lorsqu'on peut déterminer l'indice de fertilité, ou la classe de fertilité, de chaque placette avant d'ajuster le modèle. On peut imaginer cinq moyens de déterminer a priori la classe ou l'indice de fertilité.

- [1] Quand la série des mesures successives faites sur une placette inclut l'âge de référence, on connaît directement l'indice de fertilité (hauteur dominante de la placette à l'âge de référence).
- [2] Si la série de mesures n'inclut pas l'âge de référence, on peut ajuster aux données de la placette une équation comme celle de SCHUMACHER et l'utiliser pour estimer la hauteur dominante qui a été ou qui sera atteinte à l'âge de référence : on adopte cette hauteur estimée comme indice de fertilité de la placette.
- [3] A des données provenant de placettes temporaires, on peut attribuer un "niveau" de fertilité comme dans la méthode du minimum-maximum décrite au

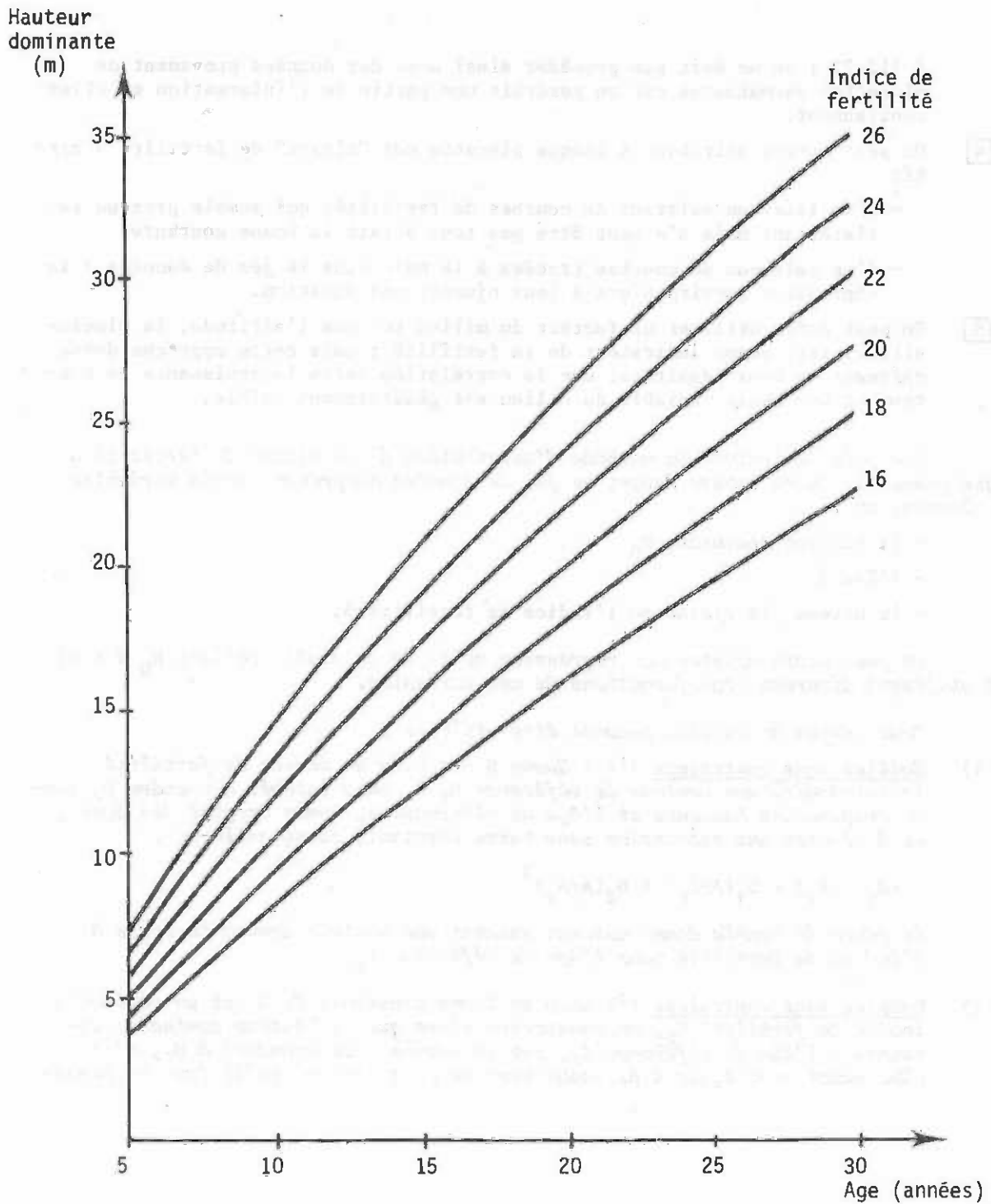


Figure 5.7

Courbes de fertilité établies à partir des données de la figure 5.6 en utilisant l'équation de Schumacher et la méthode de régression série de Bailey et Clutter.

§ 512.22 ; on ne doit pas procéder ainsi avec des données provenant de placettes permanentes car on perdrait une partie de l'information qu'elles contiennent.

4 On peut encore attribuer à chaque placette une "classe" de fertilité à partir :

- d'un faisceau existant de courbes de fertilité, qui semble presque satisfaisant mais n'a peut être pas tout à fait la bonne courbure,
- d'un faisceau de courbes tracées à la main dans le jeu de données : la régression servira alors à leur ajuster une équation.

5 On peut enfin utiliser un facteur du milieu tel que l'altitude, la pluviosité, ...etc, comme indicateur de la fertilité ; mais cette approche donne rarement de bons résultats, car la corrélation entre la croissance en hauteur et une seule variable du milieu est généralement faible.

Une fois déterminée la méthode d'attribution d'une classe de fertilité à chaque placette, on se trouve devant un jeu de données comprenant trois variables par observation :

- la hauteur dominante H_0
- l'âge A
- le niveau, la classe ou l'indice de fertilité S .

On peut alors ajuster par régression multiple un modèle reliant H_0 à A et S en utilisant diverses transformations de ces variables.

Deux sortes de modèles peuvent être utilisés :

1) Modèles sous contrainte ⁽¹⁾ : Quand S est bien un indice de fertilité (c'est-à-dire une hauteur de référence H_r), on a intérêt à prendre H_r comme origine des hauteurs et l'âge de référence A_r comme origine des âges, et à ajuster une régression sans terme constant, comme celle-ci :

$$(H_0 - H_r) = b_1(A - A_r) + b_2(A - A_r)^2$$

Ce genre de modèle donne automatiquement une hauteur dominante égale à l'indice de fertilité pour l'âge de référence A_r .

2) Modèles sans contrainte ⁽²⁾ avec un terme constant. Si S est un véritable indice de fertilité H_r , on constatera alors que la hauteur dominante atteinte à l'âge de référence A_r , sur la courbe correspondant à H_r , n'est plus exactement égale à H_r ; nous verrons plus loin ce qu'il faut en penser.

(1) en anglais = constrained models

(2) en anglais = unconstrained models

Voici deux exemples de modèles sans contrainte :

$$H_0 = b_0 + b_1A + b_2S + b_3AS + b_4A^2 \quad (13)$$

$$\log H_0 = b_0 + b_1/A + b_2S + b_3S/A \quad (14)$$

Pour S fixé, l'équation (13) représente une parabole ; la figure A22(e) de l'annexe A montre les formes qu'elle peut prendre : une seule de ces formes peut convenir à des courbes de fertilité. Toujours pour S fixé, l'équation (14) est une équation de SCHUMACHER où $k=1$ (voir figure A22(c)).

Bien d'autres modèles sont envisageables. Quel que soit le modèle choisi, il faut absolument tracer les courbes après ajustement et les superposer au nuage des données pour s'assurer qu'elles en suivent bien les tendances. Il ne faut accorder aucune confiance aux statistiques (par exemple R^2 , écart-type résiduel...) associées à la régression multiple pour décider de la validité des courbes.

Quand on travaille avec un indice de fertilité véritable (et non une simple "classe" de fertilité) en utilisant un modèle sans contrainte, on s'aperçoit en traçant les courbes obtenues que la hauteur dominante H_0 à l'âge de référence A_r n'y est généralement pas égale à l'indice de fertilité H_r : c'est que le processus d'ajustement ne comporte pas de contrainte assurant automatiquement cette coïncidence. La courbe correspondant à une valeur donnée H_r^* de l'indice de fertilité "réel" est obtenue en donnant à l'indice de fertilité "statistique" qui figure dans l'équation la valeur H_r qui fait passer la courbe par le point $[A_r, H_r^*]$.

Nous aurons ainsi, avec le modèle (13) :

$$H_r^* = b_0 + b_1A_r + b_2H_r + b_3A_rH_r + b_4A_r^2$$

soit
$$H_r = (H_r^* - b_0 - b_1A_r - b_4A_r^2) / (b_2 + b_3A_r) \quad (15)$$

et la courbe d'indice de fertilité réel H_r^* est :

$$H_0 = b_0 + b_1A + (b_2 + b_3A) \times (H_r^* - b_0 - b_1A_r - b_4A_r^2) / (b_2 + b_3A_r) + b_4A^2$$

Nous aurons ainsi, avec le modèle (14) :

$$\log H_r^* = b_0 + b_1/A_r + b_2H_r + b_3H_r/A_r$$

soit
$$H_r = (\log H_r^* - b_0 - b_1/A_r) / (b_2 + b_3/A_r) \quad (16)$$

et la courbe d'indice de fertilité réel H_r^* est :

$$\log H_0 = b_0 + b_1/A + (b_2 + b_3/A) \times (\log H_r^* - b_0 - b_1/A_r) / (b_2 + b_3/A_r)$$

La nécessité de caler les courbes de fertilité, la distinction mystérieuse et prêtant à confusion entre indice de fertilité statistique H_r et indice de fertilité réel H_r^* , ont conduit de nombreux chercheurs à utiliser plutôt des modèles sous contrainte. Malheureusement, ce n'est pas toujours possible car beaucoup de programmes standard de régression multiple n'ont pas d'option "ajustement sans terme constant".

Quand on travaille avec des classes de fertilité, le problème ne se pose pas puisqu'il n'y a pas d'obligation pour les courbes de passer par un point donné.

Les techniques de régression multiple ont l'avantage de laisser toute liberté dans le choix du modèle. Elles ont l'inconvénient de nécessiter l'estimation a priori de l'indice de fertilité, d'être statistiquement peu efficaces en ce qu'elles ignorent la nature sériée des données des placettes permanentes, et d'être fondées sur l'hypothèse fautive que la variable fertilité est connue avec une grande précision.

513 Modèles d'évaluation de la fertilité à partir des facteurs du milieu

Il est évident que les courbes de fertilité (paramétrées par un indice de fertilité ou seulement affectées d'un numéro de classe) ne peuvent servir à estimer une productivité potentielle que si on les applique à un peuplement existant. Mais, en gestion forestière, toute une catégorie de décisions suppose l'évaluation de la production potentielle d'une essence donnée avant son installation dans un milieu donné, ou très peu de temps après.

Cette évaluation est possible à deux niveaux :

- *Prévisions très globales sur la productivité maximum dans une région donnée du monde, par comparaison avec la productivité de peuplements similaires en bon état croissant ailleurs dans des conditions de milieu semblables.*
- *Prévision plus détaillée de la classe de fertilité (avec référence à un faisceau de courbes de fertilité donné pour une essence donnée) par construction, sur des peuplements de cette essence existant dans la région étudiée, d'une relation entre la classe de fertilité et certaines variables du milieu (teneurs en éléments nutritifs, profondeur et texture du sol, altitude, exposition, pluviométrie,...).*

La première approche est particulièrement adaptée aux forêts naturelles hétérogènes. En ce qui concerne les plantations monospécifiques, elle convient mieux à la phase de présélection de l'essence et de sa provenance qu'à une prévision détaillée de la production comme en traite ce manuel. Nous ne traiterons donc ici que de la deuxième approche.

513.1 Modèles fonctionnels de prévision de la classe de fertilité

Les étapes de la construction d'un tel modèle sont les suivants :

- 1 Construction de courbes de fertilité (indice ou classe) à partir de placettes permanentes d'échantillonnage dans les peuplements existants de l'essence étudiée.
- 2 Recueil, sur des placettes permanentes ou temporaires de l'essence étudiée, de données concernant le milieu ; chaque placette se voit en même temps attribuer une classe de fertilité, à partir de sa hauteur dominante et de son âge, d'après les courbes de fertilité. Pour le recueil des données concernant le milieu, on suivra les indications données au § 323.
- 3 On sélectionne et on transforme éventuellement certaines des variables du milieu (voir § 513.2) pour les introduire dans un modèle linéaire de prévision de la fertilité :

$$S = b_0 + b_1 e_1 + b_2 e_2 + \dots + b_n e_n$$

où e_1, e_2, \dots, e_n sont des transformations appropriées des variables du milieu et b_0, b_1, \dots, b_n les coefficients ajustés par régression multiple.

4

On peut déterminer l'intervalle de confiance des prévisions, comme indiqué dans l'annexe A pour les régressions multiples. Mais il faut aussi tester la valeur du modèle en comparant les classes de fertilité qu'il prédit aux classes de fertilité réelles, sur un jeu de données indépendant de celui qui a servi à l'ajuster. Si celui-ci n'a pas couvert toute la région où l'on veut appliquer le modèle, il faut que le jeu de données-test en complète la couverture.

513.2 Choix des variables du milieu à intégrer dans le modèle

Le nombre de variables du milieu mesurées peut être très grand, jusqu'à 100 sur chaque placette, alors qu'un modèle utilisable ne doit contenir qu'un petit nombre de variables faciles à mesurer.

Il y a plusieurs moyens de réduire le nombre de ces variables :

- En faisant la synthèse de certains éléments d'information : par exemple, des données climatiques peuvent être synthétisées de façon simple en "longueur de la saison humide", en "longueur de la saison de végétation" (températures supérieures à 6°C) ou en "moyenne des températures pendant la saison humide" ; elles peuvent l'être de façon plus complexe à l'aide de formules d'évapotranspiration, ...etc.
- En opérant une sélection des variables les plus significatives sur des critères statistiques. On peut utiliser pour le faire l'analyse en composantes principales ⁽¹⁾ ou, plus simplement, la régression multiple progressive ⁽²⁾.
- Par analyse graphique des variables susceptibles de représenter les principaux facteurs limitants de la croissance, en ne retenant que celles qui ont une relation évidente avec la classe de fertilité.

On n'hésitera pas à faire subir éventuellement aux variables mesurées des transformations susceptibles d'en faire des variables mieux corrélées avec la croissance. Une application évidente de ce principe concerne l'exposition ⁽³⁾, qui peut être mesurée en degrés, de 0° à 360°, les valeurs proches de 0° et de 360° correspondant à une exposition Nord : si on prend le sinus de la moitié de cet angle, on obtient alors un code exposition qui va de 0 pour les expositions Nord à 1 pour les expositions Sud, en confondant les expositions Est et Ouest.

(1) en anglais = principal components analysis

(2) en anglais = stepwise multiple regression

(3) en anglais = aspect

513.3 Problèmes posés par l'utilisation des modèles d'évaluation de la fertilité

Pour qu'un modèle d'évaluation de la fertilité soit efficace, il doit être construit sur des variables explicatives facilement mesurables. En outre, le coefficient de corrélation multiple de ces variables avec la fertilité devra être relativement élevé, de préférence supérieur à 0,80 dès qu'il y a plus de 20 observations : sinon, la relation peut être statistiquement significative et néanmoins sans utilité pratique pour la prévision en raison de son manque de précision.

Le modèle doit contenir le moins possible de variables explicatives, de préférence pas plus de trois ou quatre ; la valeur de ces variables explicatives doit être facile à déterminer en utilisant un matériel classique.

On trouvera sur la figure 5.8 un exemple de modèle de prévision de la fertilité qui répond à ces critères ; il comporte une seule variable explicative, le nombre de jours de la saison de végétation dépassant 6°C ; l'indice de fertilité y est exprimé en terme d'accroissement (annuel) moyen maximum.

Moins il y a de variables explicatives dans un modèle, plus il est robuste et susceptible de s'appliquer à une vaste région. S'il contient en revanche un grand nombre de variables explicatives, il devient très sensible aux relations qui existent entre elles et ce d'autant plus qu'elles sont fortement corrélées : des changements régionaux dans les compensations qui s'exercent entre les facteurs utilisés comme variables explicatives peuvent alors invalider un modèle trop complexe.

52 METHODES "STATIQUES" DE PREVISION DE LA PRODUCTION

Nous appellerons méthodes statiques les méthodes dans lesquelles on exprime la "production" -ou plus exactement l'état du peuplement sur pied- en fonction de l'âge, de la classe de fertilité et de l'histoire sylvicole du peuplement. Ces méthodes sont statiques en ce sens que les éclaircies ne jouent aucun rôle moteur dans le modèle et que l'on ne peut y faire varier la sylviculture que selon les quelques types de traitement bien caractérisés déjà présents dans les données.

Les caractéristiques de la production (en fait : du peuplement sur pied) qui intéressent le plus le gestionnaire sont le volume et le diamètre moyen. La détermination du volume demande la connaissance du diamètre moyen, mais aussi du nombre de tiges et de la hauteur ; on se sert généralement du nombre de tiges pour définir l'intensité des éclaircies, et de la hauteur pour déterminer la classe de fertilité.

Nous exposerons ici les méthodes statiques basées sur le diamètre moyen.

521 Méthodes graphiques de détermination de la relation diamètre moyen/hauteur dominante du peuplement

Quand on porte sur un graphique hauteur dominante (en abscisse) × diamètre moyen (en ordonnée) les données provenant d'un ensemble de placettes temporaires, on obtient un nuage de points du genre de celui de la figure 5.9(a). Si les données proviennent d'une expérience d'éclaircie, on obtiendra plutôt quelque chose qui ressemble à la figure 5.9(b), où le traitement A représente les éclaircies les plus

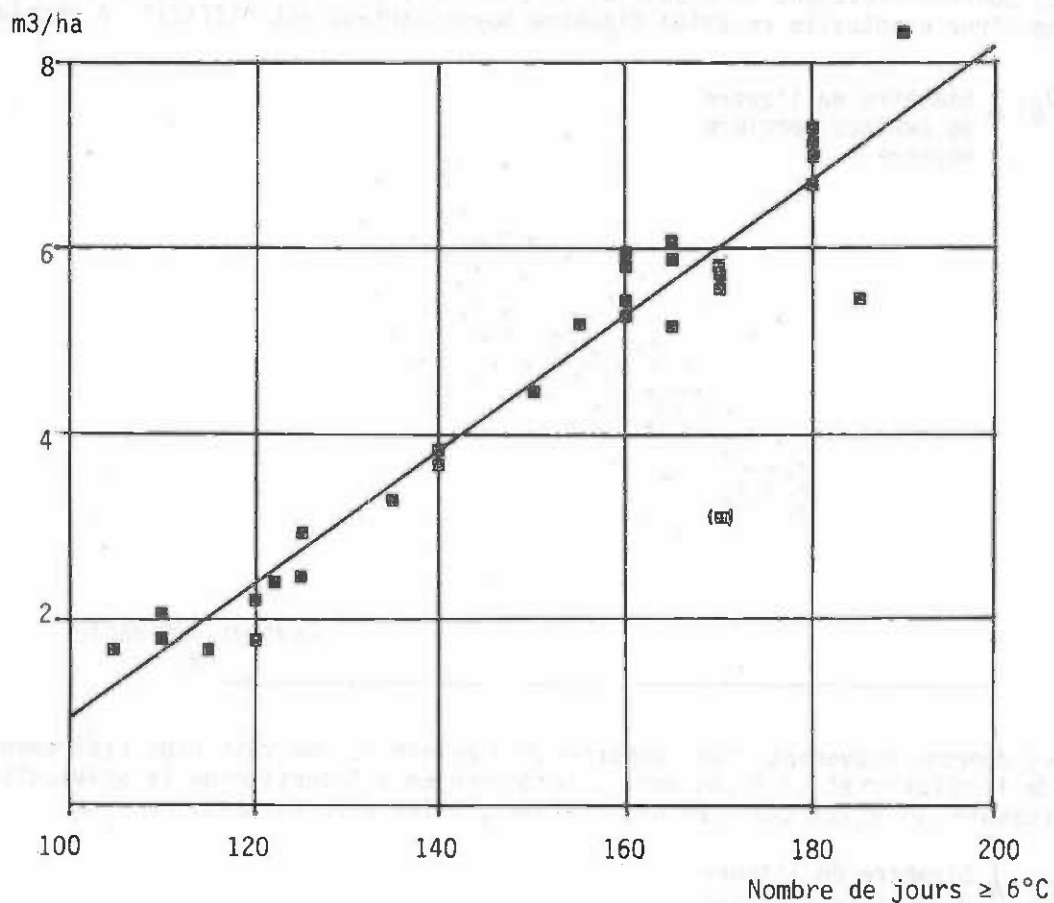


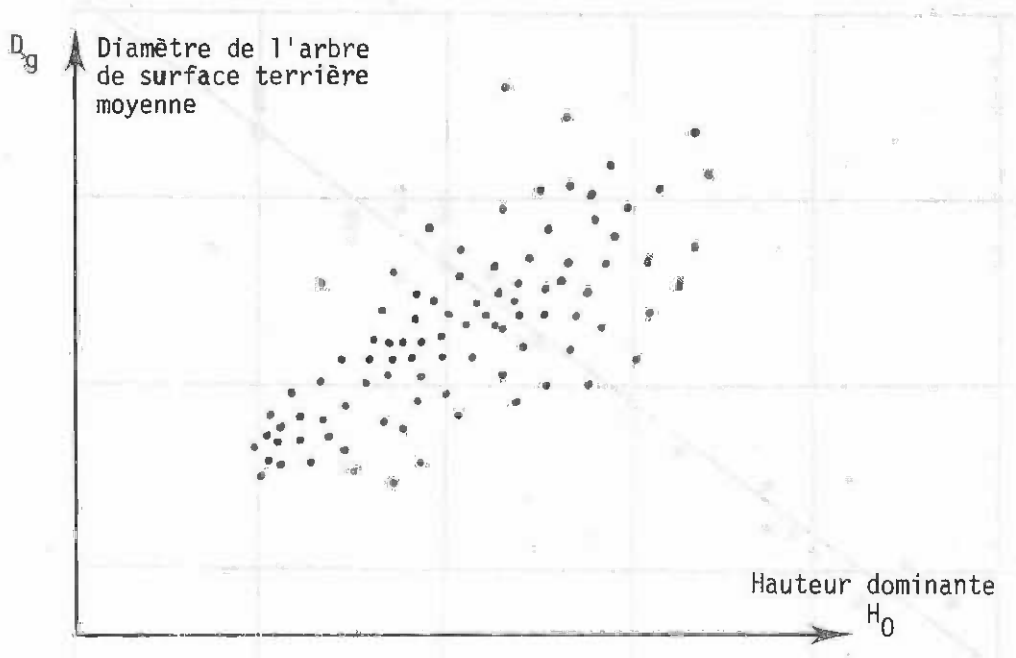
Figure 5.8

Modèle d'évaluation de la fertilité pour les peuplements de pin sylvestre (*Pinus sylvestris*) en Suède : le graphique représente la productivité moyenne des peuplements d'une région en fonction du nombre moyen annuel de jours dépassant 6°C. (d'après J. FRIES, 1978, "The assessment of growth and yield and the factors influencing it", communication spéciale au 8^e congrès forestier mondial, Djakarta).

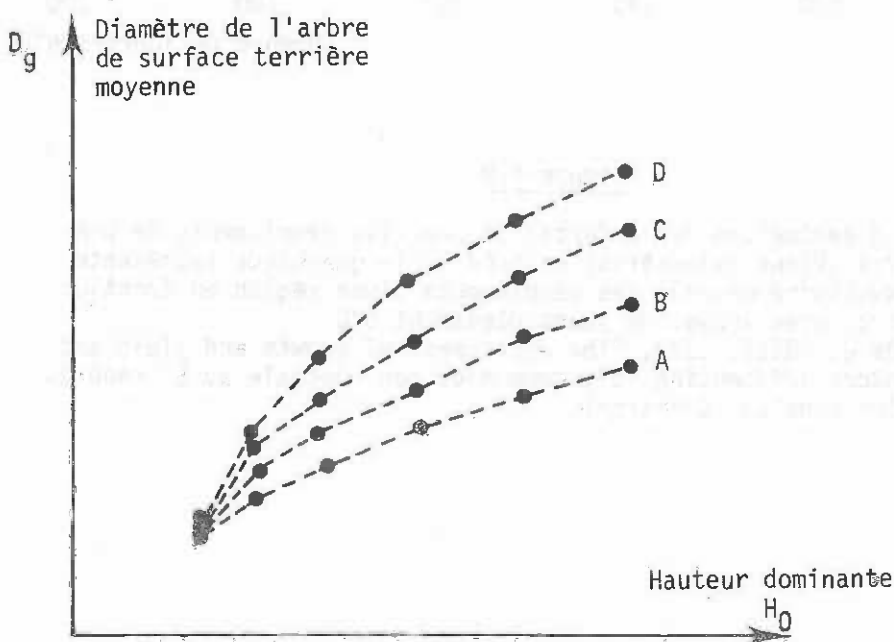
Figure 5.9

Forme de la relation entre le diamètre moyen et la hauteur dominante.

- a) Avec des données provenant de placettes temporaires, on obtient un nuage sans structure ; la forme d'une éventuelle relation diamètre moyen/hauteur est difficile à déceler.



- b) Avec des données provenant d'une expérience d'éclaircie, on voit plus clairement la forme de la relation et la façon dont elle dépend de l'intensité de la sylviculture (A représente les éclaircies les plus faibles, D les plus fortes).



légères ⁽¹⁾ (peuplement dense) et le traitement D Les éclaircies les plus fortes ⁽²⁾ (peuplement clair) ; les courbes obtenues ont généralement leur concavité tournée vers le bas.

Sur des données provenant de placettes, temporaires ou permanentes, soumises à des traitements sylvicoles qui ne diffèrent fondamentalement que par l'intensité des éclaircies ⁽³⁾, on peut tracer à la main un faisceau de courbes tout comme nous l'avons fait au § 512.1 pour les courbes de fertilité ; mais ici les différentes courbes reflèteront des différences dans l'intensité de la sylviculture.

S'il n'y a que des données de placettes temporaires, on peut définir la forme des courbes en se basant sur les enveloppes inférieure et supérieure du nuage de points : on risque évidemment de se tromper s'il existe quelques points exceptionnels ou anormaux et le résultat obtenu est de toute façon assez subjectif. Avec des données provenant de placettes permanentes, en revanche, l'allure des courbes est nettement dessinée par les lignes brisées joignant les mesures successives.

La hauteur dominante, qui figure en abscisse de ces graphiques, est un indicateur de l'âge dépendant de la fertilité : on espère ainsi intégrer implicitement l'influence de la fertilité sur la croissance en diamètre ; si cela s'avérait irréaliste, il faudrait construire des faisceaux séparés pour les différentes classes de fertilité, l'abscisse pouvant alors être indifféremment la hauteur dominante ou l'âge.

Une fois tracé le faisceau de courbes de croissance en diamètre, il reste à exprimer en termes de nombres de tiges l'intensité de la sylviculture correspondant à chaque courbe du faisceau ; nous en reparlerons au § 523.

On peut traduire en équation les courbes tracées à la main par l'une des méthodes d'approximation décrites dans l'annexe A.

522 Ajustement statistique direct de fonctions donnant le diamètre moyen

Quand on dispose de données sur le diamètre moyen, on peut leur ajuster des fonctions par les méthodes statistiques déjà décrites pour l'ajustement des courbes de croissance en hauteur (courbes de fertilité). Les variables explicatives seront la hauteur et le traitement sylvicole subi par le peuplement (repéré et désigné par T) ; on peut aussi, à la place de la hauteur dominante considérée comme apte à rendre compte de l'effet combiné de l'âge et de la fertilité, introduire séparément ces deux variables dans le modèle.

(1) en anglais = light thinning

(2) en anglais = heavy thinning

(3) Il serait en effet présomptueux d'espérer que des stratégies très différentes comme : a) éclaircies précoces - pas d'éclaircies au-delà de 20m de hauteur
b) première éclaircie à 20 m de hauteur, donnent des courbes de formes semblables. De même, des éclaircies toutes par le bas et des éclaircies toutes par le haut donneront probablement des courbes de formes assez différentes.

Dans de tels cas, il faudra vraisemblablement construire des faisceaux différents, constituant autant de variantes du modèle.

Parmi les méthodes utilisables, on peut signaler les suivantes :

(i) Avec des données provenant de placettes temporaires :

- Faire des classes de hauteur ; à l'intérieur de chacune, classer les placettes par ordre de diamètre moyen décroissant et normer ce classement pour obtenir "l'indice de traitement" de chaque placette :

$$T = i/n$$

où i est le rang de la placette et n le nombre de placettes dans la classe de hauteur. On peut alors ajuster aux données, par régression multiple, un modèle expliquant D_g en fonction de H_0 et T (voir par exemple les modèles (1) et (2) présentés plus loin).

- Ajuster une courbe moyenne par régression du diamètre moyen sur la hauteur et construire sur cette base un faisceau de courbes parallèles ou affines...

(ii) Avec des données provenant de placettes permanentes :

- Utiliser la régression simple sériée décrite au § 512.23 à propos de l'ajustement des courbes de fertilité. Il faut pour cela que l'on puisse transformer les données pour aboutir à un graphique semblable à celui de la figure 5.5 (a) (modèle à pente unique) ou 5.5(b) (modèle à constante unique); on peut essayer l'équation de SCHUMACHER ou un modèle encore plus simple du type :

$$D_g = a + bH_0^k$$

où k est un exposant compris entre 0 et 1.

- Utiliser la régression multiple sériée, soit avec la technique des variables conditionnelles décrite à l'annexe A29, soit avec des techniques plus sophistiquées de régression à plusieurs degrés.
- Utiliser la régression multiple classique, après avoir attribué à chaque placette un indice de traitement sylvicole T ; on peut y parvenir de plusieurs façons et notamment en suivant une variante de la méthode décrite ci-dessus pour les placettes temporaires : elle consiste à attribuer à une placette permanente, conduite suivant un schéma cohérent d'éclaircie, la valeur moyenne des T ainsi calculés pour les différentes mesures de la placette.

Dans les expériences d'éclaircie, on peut attribuer tout simplement à chaque traitement une valeur de $T = 1, 2, 3, \dots$ etc s'il est possible sans ambiguïté de les classer par ordre d'intensité croissante.

Une pratique courante dans l'ajustement de tels modèles est d'utiliser comme indice de traitement un facteur d'espacement ⁽¹⁾ ou même le nombre de tiges/ha sur pied. Cette méthode peut donner apparemment de bons résultats, mais elle pose des problèmes conceptuels dont nous parlerons au §523.

(¹) rapport entre l'espacement "moyen" (calculé comme si les N tiges/ha étaient plantées au carré, soit $\sqrt{10\ 000/N}$) et la hauteur dominante du peuplement.

Le choix de la fonction mathématique à ajuster sera guidé par le souci de rendre compte au mieux de la forme des courbes, qui présentent généralement une légère concavité vers le bas ; on peut dans ce cas conseiller les modèles suivants :

$$\boxed{1} \quad D_g = b_0 + b_1 H_0 + b_2 T + b_3 H_0 T + b_4 H_0^2 \quad (1)$$

$$\boxed{2} \quad D_g = b_0 + b_1 H_0^k + b_2 T + b_3 H_0^k T \quad (2)$$

On peut ajuster le second modèle par régression linéaire, pour diverses valeurs de k fixées a priori entre 0 et 1, et retenir la valeur de k qui conduit au coefficient de corrélation maximum (voir annexe A4). On peut aussi l'ajuster directement par une méthode de régression non linéaire.

Quelle que soit la fonction ajustée, il faut absolument vérifier graphiquement le résultat obtenu en le superposant aux données, de façon à pouvoir détecter tout comportement anormal.

523 Définition du traitement sylvicole en termes d'âge et de nombre de tiges

Le faisceau de courbes tracé par des méthodes graphiques (§521) ou statistiques (§522) définit des classes de traitement sylvicole. On peut alors déterminer l'évolution du nombre de tiges correspondant à chacune :

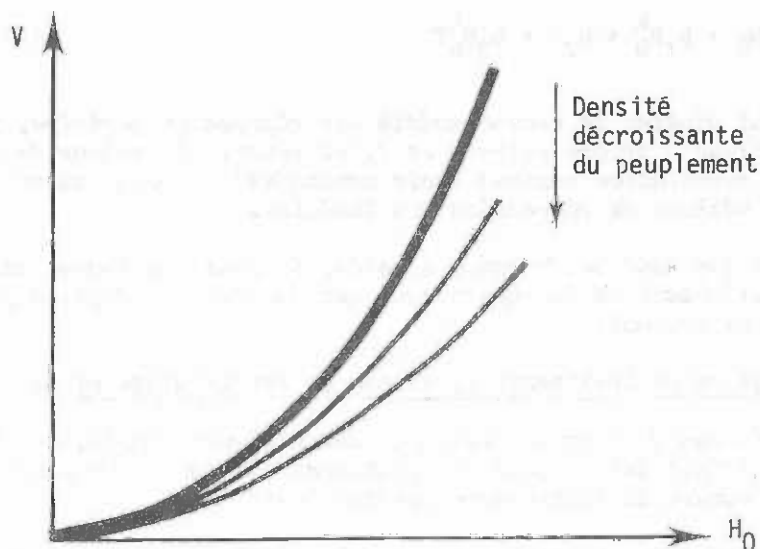
- Si la fonction ou la courbe donne le diamètre moyen en fonction de la hauteur dominante, on fait des classes de hauteur et on détermine, à partir des données, le nombre de tiges moyen dans chaque classe de hauteur, pour chaque classe de traitement ⁽¹⁾; on peut ensuite convertir, pour chaque classe de fertilité, les classes de hauteur en classes d'âge en utilisant la courbe de fertilité correspondante.
- Si la fonction donnant le diamètre moyen a été ajustée en utilisant le nombre de tiges/ha ou le facteur d'espacement comme indice de sylviculture, cette détermination est évidemment inutile. Mais la valeur des coefficients obtenus en ajustant une fonction de ce type dépend de la relation existant dans les données entre âge et nombre de tiges ; ignorer ce fait peut conduire à utiliser la fonction pour des combinaisons âge×nombre de tiges qui ne sont pas représentées dans les données, et à risquer ainsi des estimations plus ou moins fausses.

⁽¹⁾ On suppose, là encore, que l'expression du nombre de tiges N en fonction de la hauteur dominante H_0 suffit à prendre en compte implicitement, pour une classe de traitement donnée, les variations de fertilité. Si cela se révèle irréaliste, c'est-à-dire si pour une classe de sylviculture N est manifestement fonction de H_0 et de la fertilité, il faut étudier directement N par classes de traitement × âge × fertilité.

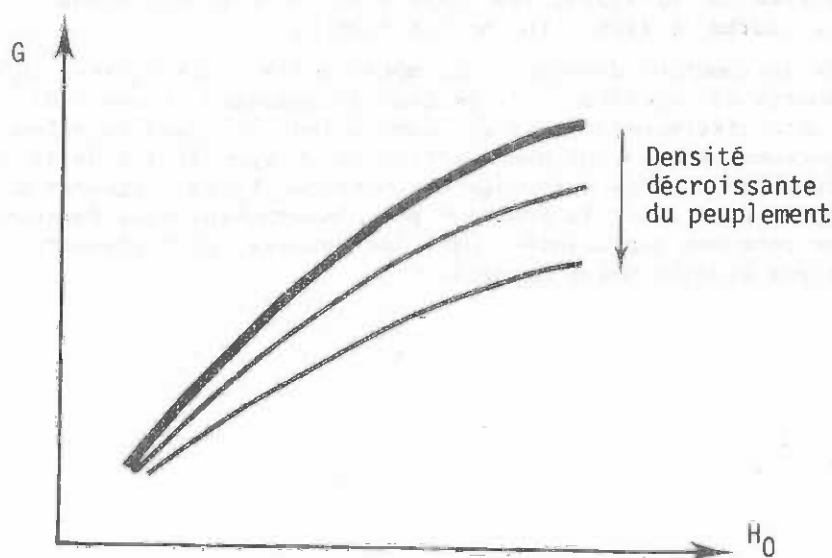
Figure 5.10

Formes des courbes volume/hauteur dominante et surface terrière/hauteur dominante.

- a) Volume par hectare en fonction de la hauteur dominante. La courbe en trait gras représente une limite qui n'est pas dépassée même avec des intensités très fortes. Si on utilise une échelle logarithmique sur les deux axes, la relation peut apparaître sous forme linéaire.



- b) Surface terrière par hectare en fonction de la hauteur dominante. La courbe en trait gras représente une limite qui n'est pas dépassée même avec des densités très fortes.



524 Modèles statiques basés sur la surface terrière ou le volume

On peut utiliser les techniques décrites ci-dessus à propos du diamètre moyen pour construire un modèle donnant la surface terrière ou le volume par hectare. Les courbes correspondantes auront l'allure indiquée sur les figures 5.10(a) et (b).

L'utilisation du diamètre moyen présente généralement des avantages en ce qui concerne la simplicité du modèle ; en outre, une partie de la réponse de la surface terrière ou du volume à des différences de densité correspond au simple effet multiplicatif de nombres de tiges différents ; ce n'est pas le cas avec le diamètre moyen, ce qui permet aux relations obtenues de mieux refléter les effets réels de la concurrence.

525 Limites des modèles statiques

Les modèles statiques de "production" ont trois gros inconvénients :

- (i) Il est peu probable que l'on obtienne un modèle cohérent et efficace si les données proviennent de peuplements qui ont subi des traitements sylvicoles très variés, voire radicalement différents.
- (ii) Une fois le modèle construit, on ne peut pas l'utiliser pour prédire l'état de peuplements soumis à des traitements sylvicoles autres que ceux représentés dans les données et intégrés dans le modèle.
- (iii) Il est difficile de déterminer la production des éclaircies si l'on ne dispose pas de l'inventaire des tiges enlevées à chaque éclaircie, ce qui est souvent le cas avec les placettes permanentes d'échantillonnage, et même avec les expériences. Dans le cas où on en dispose, la détermination du produit moyen des éclaircies par classe de traitement \times âge \times fertilité, peut permettre au modèle de donner réellement des renseignements sur la production.

Ces modèles ne peuvent être considérés comme modèles de production à part entière quand ils se contentent de décrire l'état du peuplement sur pied. En revanche, ils sont plus faciles à construire que les modèles dynamiques et peuvent être tout aussi précis, sinon plus, quand on ne leur demande que ce qu'ils sont capables de donner.

53 METHODES DYNAMIQUES DE PREVISION DE L'ACCROISSEMENT ET DE LA PRODUCTION

Un modèle est dynamique quand il modélise les variations à l'intérieur d'un système. Pour les études de production forestière, cela veut dire que le modèle est basé sur la prévision de l'accroissement courant ⁽¹⁾ en diamètre moyen, en surface terrière ou en volume.

Les modèles dynamiques présentent sur les modèles statiques l'avantage de se prêter mieux à l'expression de la véritable dépendance de cause à effet qui existe entre la densité du peuplement et les caractéristiques de la production. Par ailleurs, ils ne sont pas limités par l'obligation pour les données de représenter une série cohérente de traitements sylvicoles ; ils peuvent donc servir de cadre à l'analyse et à la synthèse de données provenant de placettes permanentes et d'expériences de types très divers.

(1) en anglais = current increment

Les modèles d'accroissement sont construits autour de fonctions exprimant l'accroissement (du diamètre moyen, de la surface terrière ou du volume) sur une courte période en fonction de la densité du peuplement (exprimée en termes de surface terrière et éventuellement d'âge), de son âge et de la fertilité. La seule hauteur dominante peut parfois suffire à rendre compte de l'effet combiné de ces deux dernières variables, comme nous l'avons supposé pour les modèles statiques.

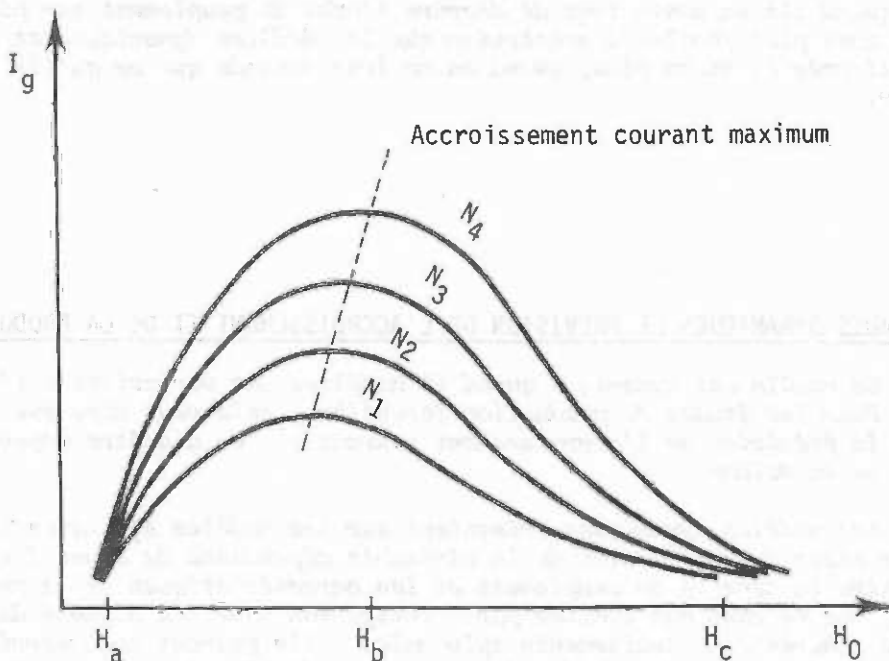
Pour faire des prévisions de production, on est amené soit à intégrer mathématiquement la fonction accroissement, soit à la sommer par itération au fil des années. Dans ce dernier cas, on écrit généralement un petit programme que nous appellerons modèle de simulation de la croissance du peuplement.

Dans les paragraphes qui suivent, nous étudions en détail ces techniques de modélisation en les appliquant à l'accroissement en surface terrière. Elles seraient peu différentes si l'on s'intéressait à l'accroissement en diamètre moyen ou en volume, mais c'est sans doute l'accroissement en surface terrière qui est le plus facile à modéliser.

531 La fonction accroissement en surface terrière

531.1 L'accroissement en surface terrière en fonction de la hauteur dominante

Si on porte sur un graphique l'accroissement courant de la surface terrière en fonction de la hauteur dominante, on obtient pour la plupart des essences en plantation des courbes ayant l'allure suivante ⁽¹⁾ :



⁽¹⁾ On suppose, là encore, que l'expression de I_g en fonction de H_0 suffit à prendre en compte implicitement, pour un niveau de densité donné, les variations de fertilité. Si cela se révèle irréaliste, il faut étudier directement I_g en fonction de la densité, de la hauteur (ou de l'âge) et de la fertilité.

où I_g est l'accroissement courant annuel en surface terrière par hectare,
 H_0 la hauteur dominante et,
 N_1, N_2, \dots des niveaux de densité croissants, exprimés par exemple en nombre de tiges à la plantation (peuplements non éclaircis).

Il peut être utile pour l'analyse des données d'y distinguer trois parties :

- [1] Entre H_a et H_b , l'accroissement augmente rapidement jusqu'à un maximum, avec des écarts importants entre les différents niveaux de densité, avant de redescendre ; pour certaines essences, cette partie de la courbe correspond à des peuplements trop jeunes pour être représentés dans les données provenant de placettes d'échantillonnage. Il est difficile de trouver un modèle de régression simple qui s'y ajuste.
- [2] Entre H_b et H_c , l'accroissement diminue rapidement avec H_0 (et donc avec le temps), avec des écarts encore bien marqués entre les différents niveaux de densité. On peut facilement modéliser cette partie de la courbe par l'une des fonctions de régression présentées sur la figure A22 de l'annexe A. Le modèle suivant, plus complexe, prend en compte l'effet du nombre de tiges :

$$\ln I_g = b_0 + b_1 H_0^{-k} + b_2 N H_0^{-k} \quad (1)$$

On peut attribuer à k une valeur a priori (1,2,...,etc) ou l'ajuster par une méthode de régression non linéaire ; on a choisi de prendre le logarithme népérien (base e) de I_g pour simplifier éventuellement l'intégration du modèle (voir §533).

- [3] Au-delà de H_c , les écarts entre niveaux de densité s'amenuisent et l'accroissement en surface terrière reste quasiment constant. Pour la plupart des essences, cette partie de la courbe n'est parcourue qu'après l'âge de l'accroissement moyen maximum en volume et ne servira donc que pour l'étude de la production de peuplements traités à longue révolution. Le modèle (1) présenté ci-dessus s'adapte également bien à cette partie des courbes d'accroissement en surface terrière.

Nous avons déjà dit qu'il était difficile de trouver une fonction linéaire unique qui s'ajuste correctement à l'ensemble de la courbe, sans biaiser l'une ou l'autre de trois parties décrites. Sachant que le point H_b se situe en général relativement tôt dans la vie du peuplement, il vaut mieux se contenter d'ajuster un modèle du type (1) ci-dessus aux données recueillies au-delà du maximum de I_g et d'estimer la surface terrière sur pied G_0 au moment où la hauteur dominante atteint H_b à l'aide d'une autre fonction à ajuster, par exemple :

$$G_0 = a + b_1 N + b_2 N^2 \quad (2)$$

où G_0 est la surface terrière initiale, correspondant à une hauteur dominante fixée H_b , et ne dépendant que du nombre de tiges alors sur pied sans considération du traitement sylvicole précédemment subi.

531.2 Autres méthodes de prévision de l'accroissement en surface terrière

On peut exprimer l'accroissement courant de la surface terrière en fonction de l'âge ou de la surface terrière sur pied. Si on utilise l'âge, il faut construire un jeu de courbes d'accroissement par classe de fertilité. Avec la surface terrière sur pied, on n'obtiendra pas de modèle satisfaisant pour des peuplements soumis à des traitements sylvicoles très variés.

Il vaut donc mieux exprimer l'accroissement en surface terrière en fonction de la hauteur dominante et de la densité, comme nous l'avons exposé ci-dessus.

531.3 Problèmes pratiqués posés par l'analyse des données d'accroissement

Les données d'accroissement présentent une grande variabilité à cause :

- de la variabilité annuelle de l'accroissement due aux fluctuations climatiques,
- des erreurs de mesure : un accroissement mesuré par différence entre deux mesures successives de précision courante peut être affecté d'une erreur du même ordre de grandeur que l'accroissement lui-même.

Il faut donc disposer d'un grand nombre de données pour obtenir une estimation correcte des coefficients de régression. En outre, comme les effets des différentes densités ne ressortent que faiblement dans la variabilité générale, il faut que des densités extrêmement différentes soient représentées dans les données si l'on veut que le modèle puisse expliciter les effets de la densité.

En définitive, on ne peut espérer obtenir un coefficient de corrélation supérieur à 0,8 ou 0,9 dans l'ajustement d'un modèle d'accroissement. Dans ces conditions, on ne peut éprouver la validité d'un modèle que sous sa forme intégrée (ou comme partie d'un modèle de simulation) en lui faisant prédire des productions finales.

En effet, la présence dans la fonction accroissement d'un petit biais qui échappe à une confrontation graphique avec les données ou à une analyse directe des résidus peut conduire, par effet cumulatif, à de grosses erreurs dans la prévision de la production.

532 Construction d'un modèle d'accroissement sur les résultats d'expériences d'espacement : hypothèse de réponse de MARSH

Il est possible, avec des techniques graphiques simples, de construire sur des expériences d'espacement (voir §241.1) des modèles capables de prédire avec précision l'accroissement de peuplements soumis à des traitements sylvicoles variés. Il faut admettre pour cela l'hypothèse de MARSH :

"Dans un type de station donné, l'accroissement d'un peuplement éclairci, ayant N tiges/ha et de surface terrière G (ou de volume V), est égal à celui d'un peuplement non éclairci :

- . ayant toujours eu N tiges/ha
- . à l'âge où il atteindrait, avec ses N tiges, la surface terrière G (ou le volume V)" ⁽¹⁾

(¹) E.K. MARSH et T.F. BURGERS, 1973, "The response of even-aged pine stands to thinning" *Forestry in South Africa*, 14, pages 103-111.

On a vérifié la validité de cette hypothèse sur plusieurs essences tropicales : elle semble conduire à une estimation précise et sans biais de l'accroissement après éclaircie, sous réserve que l'accroissement soit mesuré sur un intervalle de temps au moins égal à 3 ans.

On peut l'exploiter de la façon suivante :

- 1 Représenter graphiquement les résultats de l'expérience d'espacement (surface terrière en fonction de l'âge, voir figure 5.11), en lissant la courbe des moyennes observées pour chaque niveau de densité.
- 2 Définir le traitement sylvicole à appliquer en termes de surface terrière au moment de l'éclaircie, de nombre de tiges et de surface terrière enlevés en éclaircie.

Voici un exemple de définition d'un traitement sylvicole :

Densité de plantation : 1 300 tiges/ha

Eclaircie n°	Surface terrière au moment de l'éclaircie	G après éclaircie	N après éclaircie
1	28	22	900
2	35	28	500
3	35	30	300
CR	35	coupe rase	

- 3 Tracer sur le graphique le chemin correspondant à cette sylviculture (voir figure 5.11) et rétablir la succession réelle des âges d'éclaircie :

Eclaircie	Âges apparents	Intervalle	Âge réel au moment de l'éclaircie
1	0 - 8	8	8
2	8,5 - 13	5	12,5
3	15 - 18,5	4	16
CR	20 - 23	3	19

- 4 Reconstruire la relation surface terrière/âge, en décalant parallèlement à l'axe des âges les différents tronçons du chemin parcouru pour faire coïncider les âges après et avant chaque éclaircie. On peut alors déterminer les autres caractéristiques de la production du peuplement en fonction de l'âge, pour le type de station où a été installée l'expérience d'espacement.

Une variante de cette méthode permet de construire des courbes de production pour différentes classes de fertilité à partir d'un seul graphique, en y remplaçant en abscisse l'âge par la hauteur dominante. On fait pour cela l'hypothèse que la réponse en surface terrière des expériences d'espacement s'exprime en fonction de la hauteur d'une façon unique quelle que soit la station. Pour chaque classe de fertilité, on considère alors la hauteur dominante comme une transformation

Surface terrière
(m²/ha)

Nombre de
tiges/ha

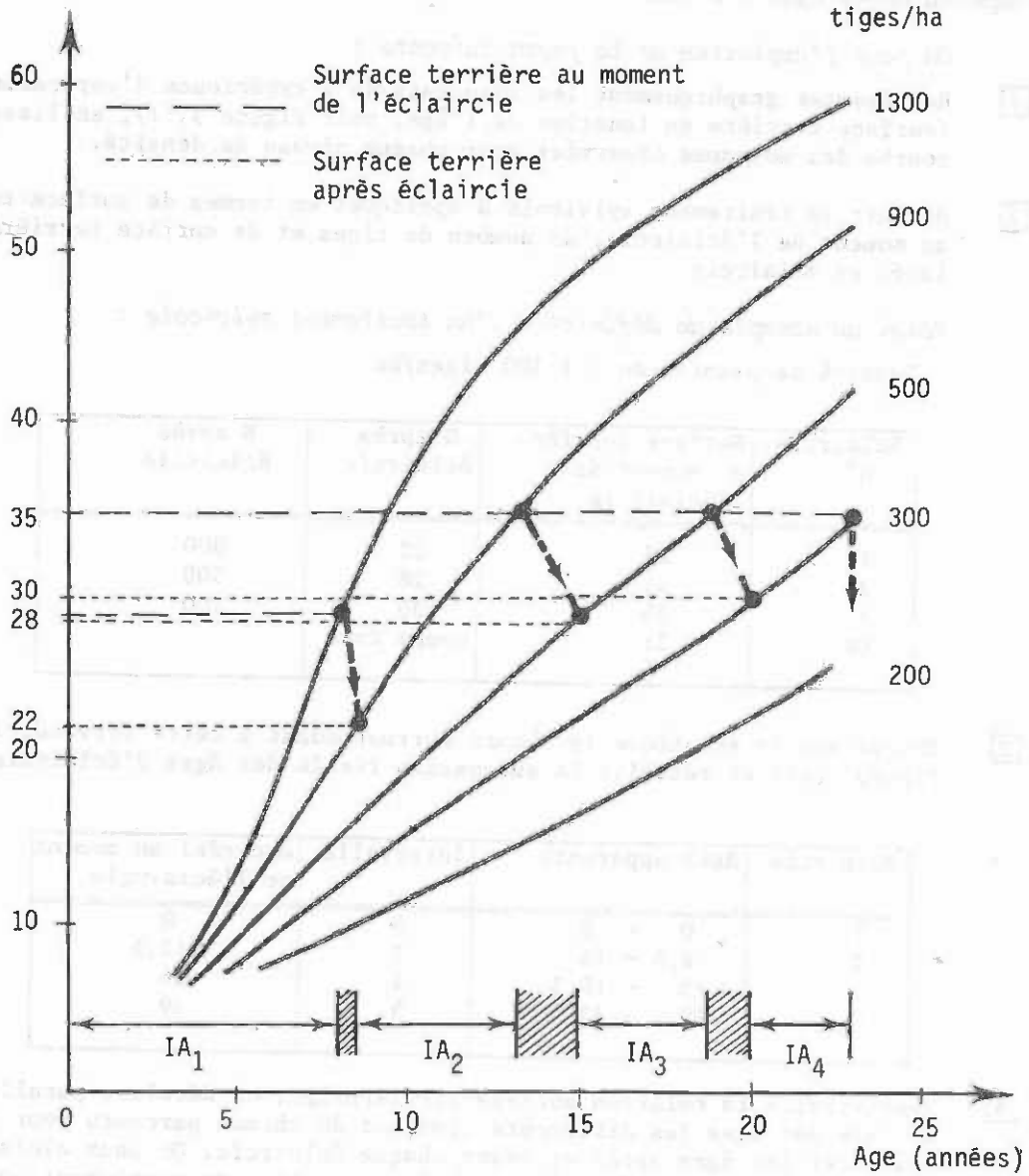


Figure 5.11

Utilisation des résultats d'une expérience d'espacement avec l'hypothèse de MARSH.

de l'âge, auquel on peut revenir en utilisant la courbe de fertilité correspondante. On peut utiliser cette méthode pour construire des tables de production hypothétiques pour des types de station non couverts par les expériences d'espacement ; mais il faut bien noter que le remplacement de combinaisons âge x fertilité variées par la seule hauteur dominante peut ne pas aboutir à une représentation suffisamment précise et réaliste des réponses liées à la station.

Si on dispose d'une fonction accroissement du genre du modèle (1) du §531.1, établie sur des placettes d'échantillonnage permanentes ou temporaires (sondages à la tarière), on peut s'en servir pour construire année par année, à nombre de tiges fixé et pour une classe de fertilité donnée, la courbe d'évolution de la surface terrière d'un peuplement non éclairci de cette densité et de cette fertilité : on peut donc obtenir ainsi, pour chaque classe de fertilité, un graphique analogue à celui de la figure 5.11, exploitable de la même façon en admettant l'hypothèse de MARSH. Ce peut être une façon d'utiliser un modèle d'accroissement existant quand on n'a pas de moyens de calcul.

533 Passage d'un modèle d'accroissement à un modèle de production par intégration

533.1 Introduction

Si l'on connaît, pour un traitement sylvicole donné, la fonction "accroissement en surface terrière (1)" d'un peuplement, son intégration sur le temps donne la fonction "production en surface terrière" de ce peuplement pour ce traitement :

$Q(t)$ = production totale (y compris les éclaircies) en surface terrière depuis l'origine du peuplement jusqu'à la date t .

L'intégration mathématique permet d'exploiter un modèle d'accroissement même si on ne dispose pas de moyens de calcul automatique, contrairement à la simulation ; malheureusement, des modèles d'accroissement apparemment simples (comme le modèle (1) du § 531.1) peuvent se révéler très difficiles à intégrer.

Cette méthode suppose de la part du chercheur des connaissances élémentaires en analyse (différentielles, intégrales, ...).

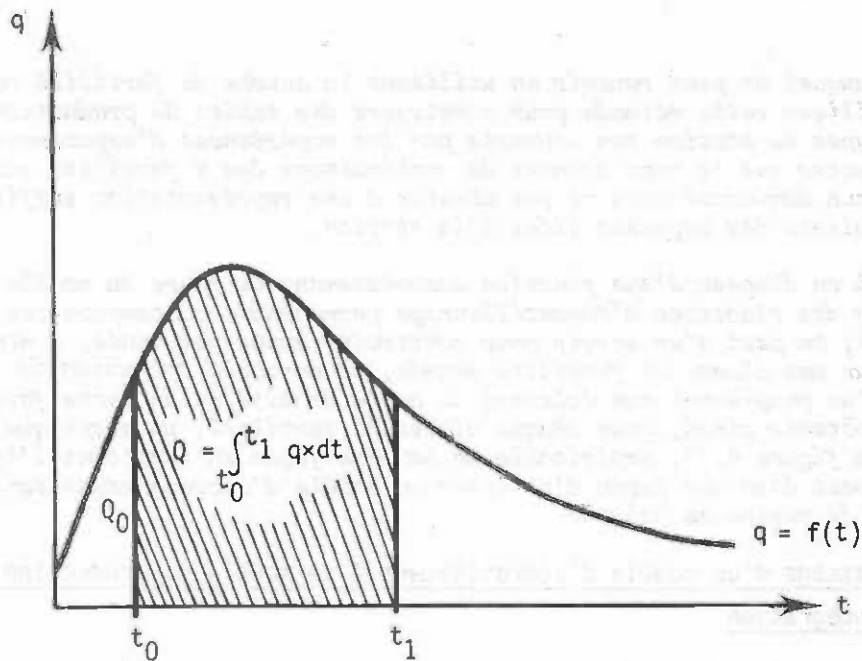
533.2 Principes de base

Soit une fonction accroissement (en surface terrière ou en volumes) :

$$q = f(t) \quad \text{où } t \text{ est le temps} \quad (1)$$

comme celle qui est représentée sur la figure ci-après :

(1) accroissement courant entre éclaircies.



La surface comprise entre l'axe des t et cette courbe jusqu'à une date t_1 est l'intégrale de q entre 0 et t, et représente la production totale depuis l'origine ⁽¹⁾ jusqu'à t_1 :

$$Q(t_1) = \int_0^{t_1} q \cdot dt \quad (2)$$

La production entre deux dates t_0 et t_1 est donnée par :

$$\int_{t_0}^{t_1} q \cdot dt = Q(t_1) - Q(t_0) \quad (3)$$

Dans un peuplement non éclairci, la courbe surface terrière/âge représente (aux arbres morts près) la fonction Q(t). Ainsi dans une station donnée, une expérience d'espacement fournira une fonction production totale ⁽²⁾, dépendante du niveau de densité N_i :

$$Q = f(t, N_i)$$

et, en dérivant par rapport au temps, la fonction accroissement correspondante :

$$q = \frac{\partial f(t, N_i)}{\partial t}$$

⁽¹⁾ en anglais = accumulated yield

⁽²⁾ en surface terrière ou en volume

Si on fait l'hypothèse que cet accroissement ne dépend que du nombre de tiges actuel, et non de l'histoire du peuplement (1), on peut l'appliquer à des peuplements éclaircis, en une démarche décrite ci-dessous qui rappelle un peu celle de MARSH.

533.3 Application d'une fonction "production totale" à une sylviculture donnée

Soit une fonction production totale, ajustée par exemple à des données provenant de peuplements non éclaircis, exprimant la surface terrière en fonction de la hauteur dominante et du niveau de densité (nombre de tiges à l'origine) :

$$G = f(H_0, N) \quad (1)$$

Si on fait une éclaircie à la date t où la hauteur dominante atteint une valeur fixée H_t , la production jusqu'à l'éclaircie est :

$$G_t = f(H_t, N_0) \quad (2)$$

où N_0 est le nombre de tiges à l'origine (après reprise),

L'éclaircie peut être définie par le nombre de tiges restant sur pied après éclaircie N'_t et :

- soit la surface terrière après éclaircie G'_t
- soit le rapport d'éclaircie r , ce qui revient au même, car :

$$r = \frac{G_t/N_0}{G'_t/N'_t} \quad (\text{voir } \S 54), \quad (3)$$

et l'hypothèse selon laquelle l'accroissement est alors la dérivée de la fonction pour $N = N'_t$ permet de déterminer la surface terrière sur pied à une date $t+i$ postérieure à l'éclaircie :

$$G_{t+i} = G'_t + f(H_{t+i}, N'_t) - f(H_t, N'_t) \quad (4)$$

533.4 Exemple d'utilisation d'une fonction "production totale"

Illustrons par un exemple ce que nous venons d'exposer. La figure 5.12 est la représentation graphique des données d'une expérience d'espacement installée dans des peuplements de *Pinus patula*, à Kwira en Tanzanie (expérience 345). Il y avait 8 traitements en placettes rectangulaires de 0,08 ha implantées systématiquement, avec 2 répétitions de chaque traitement. La figure 5.12 ne montre que les données de 4 placettes plantées respectivement à 173, 347, 694 et 1388 tiges/ha. On a ajusté à l'ensemble des données (192 observations), par régression multiple pondérée, le modèle de prévision de la surface terrière :

(1) Mais il ne faut pas se cacher qu'il s'agit d'une hypothèse assez forte, qui n'est certainement pas toujours vérifiée.

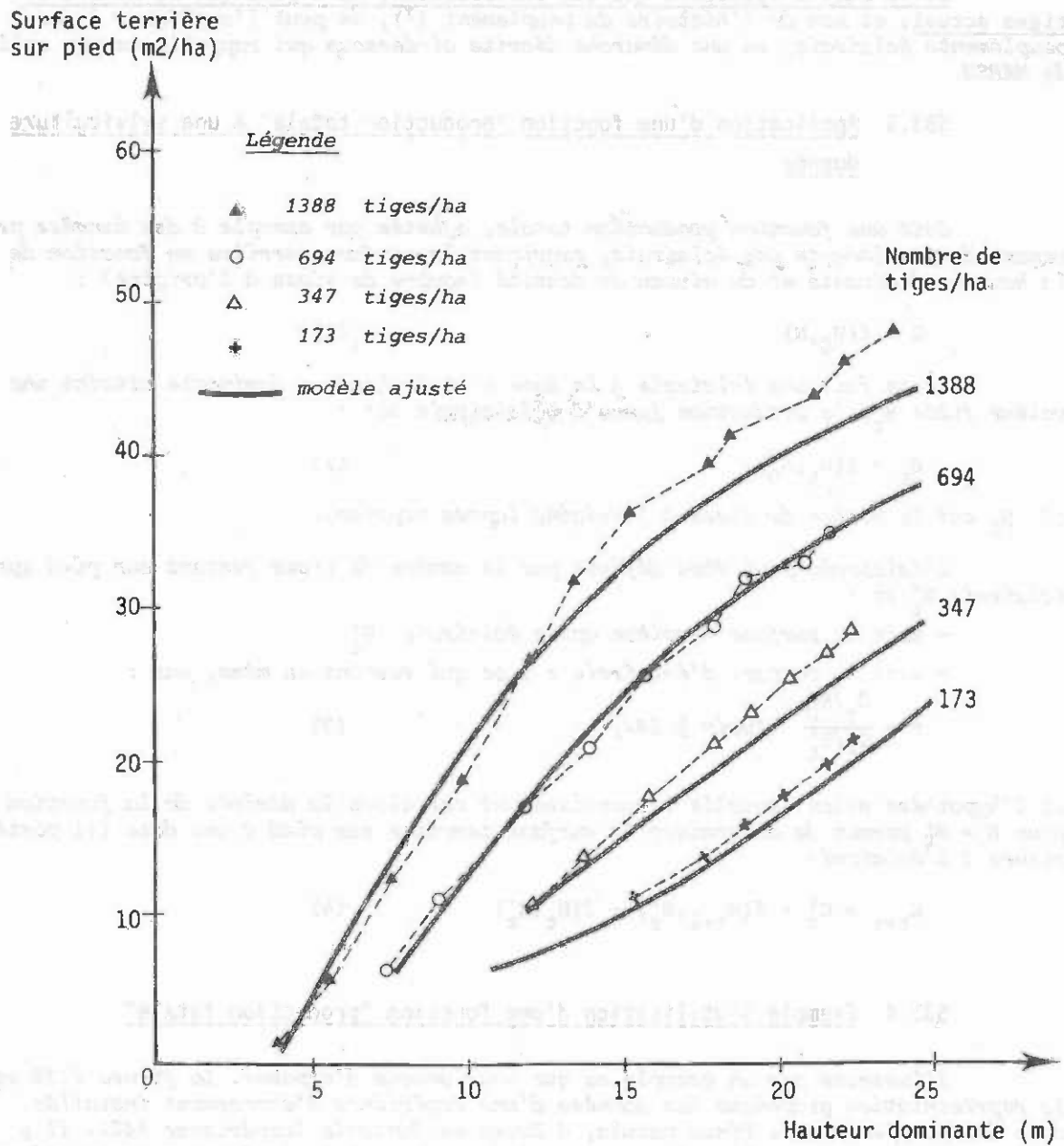


Figure 5.12

Expérience d'espacement n° 345 à Kwira (Tanzanie), dans des peuplements de *Pinus patula*.

$$\text{avec } \begin{cases} \ln G = b_0 + b_1 H^* + b_2 E + b_3 E H^* & (5) \\ H^* = 1/(H_0 - 1,30) \\ E = 100/\sqrt{N} \end{cases}$$

Les carrés des écarts étaient pondérés par G^2 pour éviter le biais négatif trop important qui résulte souvent de l'utilisation de la transformation logarithmique sur la variable expliquée. L'ajustement a donné aux coefficients les valeurs suivantes (avec $R^2=0,95$) :

$$\begin{aligned} b_0 &= 4,0865 & b_1 &= 1,5991 \\ b_2 &= 0,047838 & b_3 &= -4,3063 \end{aligned}$$

On peut voir sur la figure 5.12 ce modèle superposé aux données de 4 traitements.

On voulait, à partir de là, construire une table de production pour des peuplements plantés à 1100 tiges/ha, ramenés à 700 tiges/ha à 9 ans par une éclaircie systématique, et exploités par coupe rase à 16 ans.

Pour une fertilité moyenne, 9 ans correspondent à une hauteur dominante de 17 m, et 16 ans à 24 m. La relation (5) donne directement la surface terrière à 9 ans, avant l'éclaircie :

$$\begin{aligned} H^* &= 1/(17-1,30) = 0,0637 \\ E &= 100/\sqrt{1100} = 3,02 \\ \ln G_9 &= 4,0865 + 1,5991 \times 0,0637 + 0,047838 \times 3,02 - 4,3063 \times 3,02 \times 0,0637 \\ &= 3,5043 \\ G_9 &= e^{3,5043} = 33,26 \text{ m}^2/\text{ha} \end{aligned}$$

Puisque l'éclaircie ramenant la densité de 1100 à 700 tiges/ha est systématique, la surface terrière moyenne des arbres après éclaircie est égale à celle des arbres avant éclaircie, c'est-à-dire que $r = 1$. On a donc :

$$G'_9 = \frac{33,26/1100}{1/700} = 21,17 \text{ m}^2$$

La surface terrière enlevée en éclaircie est donc :

$$\begin{aligned} G_e &= 33,26 - 21,17 \\ &= 12,1 \text{ m}^2/\text{ha} \end{aligned}$$

L'équation (4) permet de calculer la surface terrière sur pied au moment de la coupe rase ; $f(H_{t+i}, N'_t)$ est donnée par l'équation (5) avec $H = 24$ m et

$N = 700$ tiges/ha, soit :

$$f(H_{t+i}, N'_t) = 35,98$$

$f(H_t, N'_t)$ est donné par l'équation (5) avec $H = 17$ m et $N = 700$ tiges/ha, soit :

$$f(H_t, N_t) = 26,97$$

d'où finalement la surface terrière à la coupe rase :

$$G_{CR} = 21,17 + (35,98 - 26,97) = 30,18 \text{ m}^2/\text{ha}$$

La production totale en surface terrière depuis l'origine aura été :

$$G_{tot} = 33,26 + (35,98 - 26,97) = 42,27 \text{ m}^2/\text{ha}$$

533.5 Ajustement d'une fonction "accroissement" à des données d'accroissement

Quand on ne dispose pas des données d'expériences d'espacement appropriées pour ajuster directement une fonction production totale, on peut ajuster une fonction accroissement à des données d'accroissement tirées de placettes permanentes ou temporaires, puis l'intégrer pour obtenir la fonction production totale. CLUTTER⁽¹⁾ propose un modèle d'accroissement du type :

$$I_g = G(a + bH_r - \ln G)/A \quad (6)$$

dans lequel I_g représente l'accroissement en surface terrière par hectare et par an, H_r l'indice de fertilité, a et b des coefficients qu'on peut estimer par régression linéaire en ajustant le modèle :

$$(A I_g / G + \ln G) = a + b H_r \quad (7)$$

où l'expression située à gauche représente la variable expliquée et H_r la variable explicative.

Le modèle (6) s'intègre pour donner la fonction production totale :

$$\ln G = a + b H_r - \frac{A_0}{A} (a + b H_r - \ln G_0) \quad (8)$$

où G est la production totale en surface terrière jusqu'à l'âge A , G_0 étant la production totale jusqu'à un âge "initial" A_0 . Le "niveau de densité" est indiqué ici par $\ln G_0$ ⁽²⁾, et une hypothèse analogue à celle du § 533.3 nous permettra d'utiliser cette fonction pour déterminer la surface terrière produite entre deux éclaircies de la façon suivante : si G_1 est la surface terrière après la première éclaircie faite à l'âge A_1 , et G_2 la surface terrière avant la deuxième éclaircie à l'âge A_2 , on a :

$$\ln G_2 = a + b H_r - \frac{A_1}{A_2} (a + b H_r - \ln G_1)$$

et $\Delta G_{12} = G_2 - G_1$

(1) J.L CLUTTER, 1963, "Compatible growth and yield models for Loblolly pine", Forest Science 9(3), pages 354-371

(2) et non plus par un nombre de tiges

534 Utilisation d'un modèle d'accroissement en simulation

534.1 Eléments nécessaires à la construction d'un modèle de simulation

Dans la majorité des cas, il est beaucoup plus facile, pour utiliser un modèle d'accroissement, de passer par un modèle de simulation de la croissance que de procéder par intégration mathématique, mais cela implique que l'on ait accès à un petit ordinateur. Si l'on veut procéder par simulation, il faut avoir déterminé, à partir de placettes permanentes ou de données expérimentales, les fonctions suivantes :

- [1] La fonction hauteur dominante/âge et indice de fertilité (courbes de fertilité) :

$$H_0 = f_1(A, H_r)$$

- [2] La fonction accroissement en surface terrière, où cet accroissement I_g est exprimé en fonction de :

- la hauteur dominante (ou l'âge et l'indice de fertilité)

- la surface terrière et/ou le nombre de tiges par hectare

$$\text{soit : } I_g = f_2(H_0, N) \quad \text{ou} \quad I_g = f_2(H_0, G)$$

$$\text{ou } I_g = f_2(A, N, H_r) \quad \text{ou} \quad I_g = f_2(A, G, H_r) \dots \text{etc.}$$

Le modèle de CLUTTER présenté au § 533.5 est un exemple de ce dernier type de fonction accroissement.

- [3] Le rapport entre la surface terrière moyenne des arbres avant éclaircie et la surface terrière moyenne des arbres restant sur pied après éclaircie (rapport d'éclaircie, voir § 541), à définir pour chaque éclaircie ou chaque type d'éclaircie

$$\frac{G/N}{G'/N'} = f_3$$

- [4] La fonction volume, exprimant le volume en fonction de la hauteur et de la surface terrière :

$$V = f_4(H_0, G)$$

- [5] La surface terrière initiale du peuplement à une hauteur ou à un âge initial A_0 marquant le point de départ de la simulation, en fonction du nombre de tiges à la plantation :

$$G_0 = f_5(N_0)$$

Ces fonctions n'ont pas nécessairement une expression mathématique, mais peuvent être tabulées ; elles peuvent aussi offrir un jeu de valeurs ou de fonctions dont l'une sera sélectionnée sur un test lors de la simulation.

534.2 Méthode de construction d'un modèle de simulation

Pour construire un modèle de simulation, on transcrit les relations fonctionnelles ainsi que les liens logiques qui existent entre elles sous forme d'instructions ⁽¹⁾ d'un langage de programmation évolué comme le FORTRAN, le BASIC, l'ALGOL, ...etc. L'ensemble de ces instructions constitue un programme. Ce programme en "langage source" est "compilé" par l'ordinateur qui le traduit dans son propre langage interne et peut alors en exécuter les instructions : le modèle est prêt à fonctionner.

Au début de l'exécution, il faudra lui fournir certains renseignements définissant le peuplement à simuler :

- l'indice de fertilité et le nombre de tiges initial du peuplement,
- la date, l'intensité et le rapport d'éclaircie, pour chaque éclaircie.

Le programme fournira alors ses résultats, sous forme de tableau ou de graphiques selon les instructions que l'on y a mises.

On peut distinguer deux phases dans le déroulement logique d'une simulation:

- Une phase d'initialisation, qui comprend la définition des valeurs initiales de la surface terrière, du nombre de tiges et de la hauteur du peuplement, ainsi que l'impression des titres,
- Une phase dynamique, au cours de laquelle on calcule l'accroissement sur une période de n années prise comme unité de temps ($1 \leq n \leq 5$), on l'ajoute au matériel existant, on "réalise" l'éclaircie éventuellement prévue, on imprime l'état du matériel sur pied, et on recommence après avoir ajouté n à l'âge du peuplement. Cette phase dynamique s'arrête à la coupe rase, ou à un âge limite déterminé.

534.3 Exemple de modèle de simulation simple

Nous construirons dans cet exemple un modèle de simulation simple pour des peuplements équiennes de *Pinus patula*, et nous l'écrirons en langage BASIC.

Les fonctions du modèle sont les suivantes :

- [1] La fonction hauteur dominante/âge est empruntée à ALDER ⁽²⁾ :

$$\ln H_0 = b_0 + b_1/A + ((\ln H_r - c_1)/c_2) \times (b_2 + b_3/A)$$

où les valeurs de b_0 , b_1 , b_2 , b_3 ont été obtenues par régression à par-

⁽¹⁾ en anglais = statements

⁽²⁾ D. ALDER, 1977, "A growth and management model for coniferous plantations in East-Africa", D. Phil. thesis, Oxford University

Tableau 5.1

Valeurs de l'accroissement en surface terrière en fonction de la densité et de la hauteur dominante, observées dans l'expérience 345 de Kwira (Tanzanie).

Le tableau donne, en m³/ha/an, la valeur de l'accroissement en surface terrière par classe de hauteur dominante et d'espacement calculé comme dans une plantation au carré ($e = \sqrt{10000/N}$).

Classe de hauteur dominante (en m)	Classe d'espacement						
	(en m)						
	2-3	3-4	4-5	5-6	6-7	7-8	8-9
3 - 5	3,89	2,87					
5 - 7	6,59	3,69					
7 - 9	7,22	4,49					
9 - 11	7,33	6,10	3,83				
11 - 13	5,90	5,39	4,55	3,56	3,17		
13 - 15	4,59	4,59	4,41	3,50	2,84	1,91	
15 - 17	2,82	3,30	4,35	2,93	2,34	2,18	1,51
17 - 19	2,24	2,22	3,34	2,50	2,00	2,11	1,54
19 - 21	2,09	2,03	2,40	1,73	1,42	1,72	1,43
21 - 23	2,01	1,57	1,99	1,32	1,28	1,37	1,27
23 - 25	1,86	1,93	1,59				

tir de données provenant de placettes permanentes d'échantillonnage :

$$\begin{aligned} b_0 &= 3,6068 & b_1 &= - 17,513 \\ b_2 &= 0,008057 & b_3 &= 0,3308 \end{aligned}$$

et où :

$$c_1 = b_0 + b_1/15 \quad (\text{l'âge de référence pour } H_T \text{ est 15 ans})$$

$$c_2 = b_2 + b_3/15$$

2 La fonction accroissement en surface terrière provient de l'expérience 345 de Kwira (Tansanie) : elle est représentée par le tableau 5.1, obtenu en faisant la moyenne des accroissements observés par classe de hauteur dominante et d'espacement initial, et utilisée telle quelle dans le programme.

3 Les rapports d'éclaircie sont ainsi définis :

Eclaircies systématiques : $r = 1$

Eclaircies sélectives : $r = 0,7$ pour la première éclaircie
 $r = 0,9$ pour les suivantes.

4 La fonction volume est le tarif de cubage de peuplement pour le Pinus patula au Kenya :

$$V_t = N \times (-0,0072 + 0,00002887 \times D_g^2 + 0,00002077 \times D_g H_0 + 0,000032765 \times D_g^2 H_0)$$

qui donne le volume total. Le volume marchand à la découpe 20 cm de diamètre est donné par la relation :

$$V_m = V_t \times (0,97352 - 21,9737 \times \exp(-0,15407 D_g))$$

5 La surface terrière est initialisée à 0 à l'âge de 5 ans, quelle que soit la densité de plantation.

Voici le listage du programme correspondant, en BASIC :

```
10 REM EXEMPLE DE MODELE DE SIMULATION DE LA CROISSANCE D'UN PEUPEMENT
20 PRINTER IS 7,6,WIDTH(132)
30 OPTION BASE 1
40 REM
50 REM PHASE D'INITIALISATION
60 PRINT "MODELE DE SIMULATION POUR DES PLANTATIONS DE PINUS PATULA"
70 PRINT "INDIQUEZ: NOMBRE DE TIGES INITIAL, INDICE DE FERTILITE"
80 INPUT N,F
90 PRINT N,F
100 REM LECTURE DE LA TABLE DES ACCROISSEMENTS EN SURFACE TERRIERE SELON
110 REM LA CLASSE D'ESPACEMENT (1ER INDICE) ET DE HAUTEUR (2E INDICE)
120 DIM T(7,11)
130 MAT READ T
140 DATA 3.89,6.59,7.22,7.33,5.90,4.59,2.82,2.24,2.09,2.01,1.86
150 DATA 2.87,3.69,4.49,6.10,5.39,4.59,3.30,2.22,2.03,1.57,1.93
160 DATA 0 ,0 ,0 ,3.83,4.55,4.41,4.35,3.34,2.40,1.99,1.59
```

```
170 DATA 0 ,0 ,0 ,0 ,3.56,3.50,2.93,2.50,1.73,1.32,0
180 DATA 0 ,0 ,0 ,0 ,3.17,2.84,2.34,2.00,1.42,1.28,0
190 DATA 0 ,0 ,0 ,0 ,0 ,1.91,2.18,2.11,1.72,1.37,0
200 DATA 0 ,0 ,0 ,0 ,0 ,0 ,1.51,1.54,1.43,1.27,0
210 REM LECTURE DES COEFFICIENTS DE LA RELATION HAUTEUR/AGE
220 READ B0,B1,B2,B3
230 DATA 3.6068,-17.513,0.008057,0.3308
240 LET C1=B0+B1/15
250 LET C2=B2+B3/15
260 REM INITIALISATION A ZERO DE LA SURFACE TERRIERE A L'AGE DE 5 ANS
270 LET G=0
280 LET A=5
290 REM LECTURE DES SPECIFICATIONS DE LA SYLVICULTURE
300 PRINT "AGE D'EXPLOITABILITE DU PEUPEMENT?"
310 INPUT R
320 PRINT R
330 DIM A1(10),N1(10),T$(10)
340 PRINT "NOMBRE D'ECLAIRCIES?"
350 INPUT M
360 PRINT M
370 IF M=0 THEN 430
380 FOR I=1 TO M
390 PRINT "INDIQUEZ: AGE,NB DE TIGES LAISSE SUR PIED,TYPE('SYS'OU'SEL') DE L'E
CLAIRCIE ",I
400 INPUT A1(I),N1(I),T$(I)
410 PRINT A1(I),N1(I),T$(I)
420 NEXT I
430 LET M=1
440 REM IMPRESSION DES TITRES
450 PRINT
460 PRINT "PEUPEMENT SUR PIED APRES ECLAIRCIE EVENTUELLE";TAB(80);"ECLAIRCIES
.
470 PRINT
480 PRINT " AGE HDOM TIGES/HA DIAM(G) ST/HA VOL/HA VOL/HA(20CM)
TIGES/HA DIAM(G) ST/HA VOL/HA VOL/HA(20CM)"
490 REM
500 REM PHASE DYNAMIQUE DE LA SIMULATION
510 REM CALCUL DE LA HAUTEUR DOMINANTE
520 LET H=EXP(B0+B1/A+(LOG(F)-C1)/C2*(B2+B3/A))
530 REM CLASSES DE HAUTEUR ET DE DENSITE POUR LA DETERMINATION DE
540 REM L'ACCROISSEMENT EN SURFACE TERRIERE
550 LET H1=(H-2)/2
560 IF H1<12 THEN 580
570 LET H1=11
580 LET E1=SQR(1/N)*100-1.5
590 IF T(E1,H1)=0 THEN 1140
600 REM ADDITION DE L'ACCROISSEMENT A LA SURFACE TERRIERE COURANTE
610 LET G=G+T(E1,H1)
620 REM TEST:UNE ECLAIRCIE EST-ELLE PREVUE DANS L'ANNEE EN COURS?
630 IF A=A1(N) THEN 670
640 REM CALCUL DU DIAMETRE DE L'ARBRE DE S.T. MOYENNE ET DES VOLUMES
650 LET D=SQR(G/(N*.00007854))
660 LET V=N*(-.0072+.00002887*D^2+.00002077*D*N+.00003276*D^2*N)
661 IF V>0 THEN 670
662 LET V=0
670 LET V1=V*(.97352-21.9737*EXP(-.15407*D))
680 IF V1>0 THEN 700
690 LET V1=0
700 REM IMPRESSION DES CARACTERISTIQUES DU PEUPEMENT SUR PIED
710 REM APRES ECLAIRCIE EVENTUELLE
720 PRINT USING 730;A,H,N,D,G,V,V1
730 IMAGE #,4D,4D.D,8D,8D.D,3(6D.D)
740 REM IMPRESSION DES CARACTERISTIQUES DE L'ECLAIRCIE EVENTUELLE
750 IF T2=0 THEN 780
```

```
760 PRINT USING 770;N2,D1,G1,V2,V3
770 IMAGE #,12X,9D,7D.D,7D.D,6D.D,7D.D
780 REM ON AJOUTE 1 A L'AGE ET ON REPETE LA PHASE DYNAMIQUE
790 PRINT
800 LET A=A+1
810 LET T2=0
820 IF A<R THEN 500
830 PRINT
840 PRINT TAB(20);"FIN DE LA SIMULATION"
850 STOP
860 REM
870 REM SIMULATION DE L'ECLAIRCIE SYSTEMATIQUE OU SELECTIVE
880 REM DETERMINATION DU RAPPORT D'ECLAIRCIE A PARTIR DU TYPE DE L'ECLAIRCIE
890 IF T*(M)="SEL" THEN 920
900 LET T3=1
910 GOTO 960
920 IF M>1 THEN 950
930 LET T3=.7
940 GOTO 960
950 LET T3=.9
960 REM CALCUL DU NB DE TIGES ET DE LA SURFACE TERRIERE ENLEVES N2 ET G1
970 LET N2=N-N1(M)
980 LET G2=N1(M)/N/(T3/G)
990 LET G1=G-G2
1000 REM NB DE TIGES ET S.T. DU PEUPEMENT RESTANT SUR PIED
1010 LET G=G2
1020 LET N=N1(M)
1030 REM CALCUL DU DIAMETRE MOYEN ET DES VOLUMES DE L'ECLAIRCIE
1040 LET D1=SQR(G1/(N2*.0007854))
1050 LET V2=N2*(-.0072+.00002887*D1^2+.00002077*D1^3+.00003276*D1^4)
1060 LET V3=V2*(.97352-21.9737*EXP(-.15407*D1))
1070 IF V3=0 THEN 1090
1080 LET V3=0
1090 REM ON AJOUTE 1 AU NUMERO D'ECLAIRCIE ET ON RETOURNE A LA PARTIE
1100 REM PRINCIPALE DE LA SIMULATION
1110 LET M=M+1
1120 LET T2=1
1130 GOTO 640
1140 REM FIN DU PROGRAMME QUAND LA SIMULATION SORT DU DOMAINE OU L'ON
1150 REM DISPOSE DE DONNEES D'ACCROISSEMENT
1160 PRINT TAB(20);"LA SIMULATION SORT DU DOMAINE DE LA FONCTION D'ACCROISSEMENT
T"
1170 PRINT TAB(20);"ARRET DU PROGRAMME"
1180 END
```

La figure 5.13 montre les résultats fournis par une exécution du programme.

Le tableau 5.2 donne la définition de chacune des variables utilisées dans le programme.

Le but de cet exemple est de montrer la souplesse de la technique de simulation ; il ne faudrait pas en conclure que les fonctions qui y sont présentées sont les meilleures que l'on puisse trouver ...

Figure 5.13

Edition des résultats produits par un modèle
de simulation de la croissance

MODELE DE SIMULATION POUR DES PLANTATIONS DE PINUS PATULA

INDIQUEZ: NOMBRE DE TIGES INITIAL, INDICE DE FERTILITE

1700 23

AGE D'EXPLOITABILITE DU PEUPELEMENT?

16

NOMBRE D'ECLAIRCIES?

1

INDIQUEZ: AGE, NB DE TIGES LAISSE SUR PIED, TYPE ('SYS' OU 'SEL') DE L'ECLAIRCIE 1

8 700 SYS

PEUPELEMENT SUR PIED APRES ECLAIRCIE EVENTUELLE

AGE	HDOM	TIGES/HA	DIAM(G)	ST/HA	VOL/HA	VOL/HA(20CM)
5	6.2	1700	7.0	6.6	8.7	0.0
6	8.6	1700	10.2	13.8	45.3	0.0
7	10.8	1700	12.6	21.1	96.0	0.0
8	12.9	700	14.2	11.1	61.8	0.0
9	14.8	700	16.9	15.7	101.7	0.0
10	16.6	700	18.6	19.0	137.8	0.0
11	18.1	700	19.7	21.2	168.4	0.0
12	19.5	700	20.6	23.3	198.8	10.0
13	20.8	700	21.5	25.3	230.1	38.5
14	21.9	700	22.1	26.9	257.9	63.1
15	23.0	700	22.7	28.4	285.9	89.5
16	24.0	700	23.5	30.4	317.9	122.7

ECLAIRCIES

TIGES/HA	DIAM(G)	ST/HA	VOL/HA	VOL/HA(20CM)
1000	14.2	15.9	87.0	0.0

FIN DE LA SIMULATION

Tableau 5.2

Définition des variables utilisées dans le modèle de simulation de la croissance.

Variable	Dimension	Description de l'utilisation
A		Age de peuplement, en années.
A ₁	(10)	Liste des âges des éclaircies.
B ₀ B ₁ B ₂ B ₃		Coefficients b ₀ , b ₁ , b ₂ , b ₃ de la fonction croissance en hauteur.
C ₁ C ₂		Coefficients c ₁ , c ₂ de la fonction croissance en hauteur.
D		Diamètre de l'arbre de surface terrière moyenne du peuplement sur pied.
D ₁		Diamètre de l'arbre de surface terrière moyenne enlevé en éclaircie.
E ₁		Classe de densité du peuplement, variant de 1 pour des espacements de 2-3m à 7 pour des espacements de 8m et plus.
F		Indice de fertilité du peuplement, en mètres de hauteur dominante à 15 ans.
G		Surface terrière/ha du peuplement sur pied.
G ₁		Surface terrière/ha des arbres enlevés en éclaircie.
G ₂		Surface terrière/ha du peuplement sur pied immédiatement après éclaircie.
H		Hauteur dominante du peuplement.
H ₁		Classe de hauteur dominante variant de 1 pour des hauteurs de 3-5m, à 11 pour des hauteurs de plus de 23m.
I		Compteur : numéro de l'éclaircie pendant la lecture des paramètres qui la décrivent.
M		Au début = nombre d'éclaircies. Pendant la phase dynamique = numéro de la prochaine éclaircie à intervenir.
N		Nombre de tiges/ha du peuplement sur pied.
N ₁	(10)	Liste des nombres de tiges/ha à laisser sur pied après les éclaircies.
N ₂		Nombre de tiges/ha enlevés au cours d'une éclaircie.
R		Age d'exploitabilité du peuplement en années.
T	(7,11)	Tableau des accroissements en surface terrière correspondant à une classe d'espacement (1ère dimension) et de hauteur dominante (2ème dimension). Ce sont les valeurs du tableau 5.1.
T ₂		Indicateur = 1 si le peuplement est éclairci pendant l'année en cours de simulation, = 0 sinon ; il contrôle le format d'impression.
T ₃		Rapport d'éclaircie comme défini dans le texte pour chaque type d'éclaircie (= 1 pour une éclaircie systématique, 0,7 pour la 1ère éclaircie sélective, 0,9 pour les autres éclaircies sélectives).
T ₄	(10)	Liste des types des éclaircies (pour chaque éclaircie : 'SYS' = éclaircie systématique, 'SEL' = éclaircie sélective).
V		Volume total sur écorce du peuplement sur pied, à l'hectare.
V ₁		Volume sur écorce à la découpe 20cm de diamètre du peuplement sur pied, à l'hectare.
V ₂		Volume/ha total enlevé en éclaircie.
V ₃		Volume/ha à la découpe 20 cm enlevé en éclaircie.

54 L'ECLAIRCIE

Dans les peuplements équiennes gérés pour la production de bois d'oeuvre, une partie de cette production est récoltée par des éclaircies commercialisables ; celles-ci peuvent être précédées d'éclaircies non commercialisables, destinées à réduire précocement la densité du peuplement et à favoriser d'autant l'augmentation du diamètre moyen pour les éclaircies suivantes et la récolte finale.

De toute évidence, la modélisation des éclaircies est au coeur de toute tentative de modélisation dynamique de la production. On peut en gros caractériser une éclaircie par le nombre de tiges et la surface terrière enlevés à l'hectare, du moins pour les modèles de peuplement dont nous traitons ici (ce n'est plus vrai pour les modèles individuels d'arbre, où il faut décrire la distribution des arbres enlevés).

541 Le rapport d'éclaircie

Une caractéristique importante des éclaircies, leur type ⁽¹⁾ (caractère plus ou moins "par le haut" ou "par le bas"), peut être définie par le rapport d'éclaircie ⁽²⁾ :

$$\begin{aligned} r = \text{rapport d'éclaircie} &= \frac{\text{Surface terrière moyenne des arbres avant éclaircie}}{\text{Surface terrière moyenne des arbres après éclaircie}} \\ &= \frac{G/N}{G'/N'} \\ &= \left(\frac{\text{diamètre moyen avant éclaircie}}{\text{diamètre moyen après éclaircie}} \right)^2 \end{aligned}$$

⁽¹⁾ en anglais = type of thinning

⁽²⁾ en anglais = thinning ratio

Certains utilisent un autre rapport d'éclaircie, défini de la façon suivante :

$$r' = \frac{\text{surface terrière moyenne des arbres enlevés en éclaircie}}{\text{surface terrière moyenne des arbres avant éclaircie}}$$

Ce deuxième rapport, plus facile à interpréter et plus parlant, a aussi l'avantage :

- de mieux refléter le type de l'éclaircie à proprement parler, indépendamment de son intensité,
- d'être plus sensible, variant par exemple de 0,2 à 1,2 dans l'échelle des types d'éclaircie représentée dans le tableau ci-après.

Il rend exactement les mêmes services que r dans un modèle de production,

On peut, pour fixer les idées, indiquer les ordres de grandeur suivants :

Type d'éclaircie	Intensité (indiquée par le rapport N_e/N)	Valeur du rapport d'éclaircie r
Mortalité naturelle sur une courte période	3 %	0,98
Eclaircie précoce très "par le bas"	20 % (faible)	0,85
	60 % (forte)	0,65
Eclaircies ultérieures "par le bas"	20 % (faible)	0,90
	50 % (forte)	0,80
Eclaircie "mixte" intervenant aussi dans l'étage dominant	25 %	0,95
Eclaircie systématique	quelconque	1,00
Eclaircie uniquement dans l'étage dominant	15 %	1,05

On peut déterminer empiriquement le rapport d'éclaircie par l'observation, sur des placettes temporaires :

- d'éclaircies réellement exploitées par des entrepreneurs,
- d'éclaircies simulées (marquées, mais non exploitées).

La première méthode fournit une évaluation plus réaliste de ce qui est pratiqué, mais la seconde convient mieux à la détermination des caractéristiques d'éclaircies hypothétiques que l'on souhaite tester sur un modèle.

D'une façon générale, il faut éviter d'évaluer sur des placettes permanentes les caractéristiques de la sylviculture pratiquée : ces placettes sont généralement délimitées de façon très visible, et l'on risque donc de leur faire subir, plus ou moins consciemment, un traitement particulier lors des éclaircies.

542 Estimation de la production enlevée en éclaircie (modèles statiques)

Pour estimer la production enlevée en éclaircie dans un modèle statique donné, donc correspondant à un traitement sylvicole bien défini, il faut d'abord savoir si la fonction décrivant l'état du peuplement s'appuie sur des observations faites immédiatement avant éclaircie, immédiatement après, ou si, comme c'est presque toujours le cas, il s'agit d'un mélange non contrôlé d'observations faites avant, après et entre les éclaircies.

Supposons que nous soyons dans ce dernier cas, que le modèle statique de "production" décrive l'évolution du diamètre moyen du peuplement sur pied, et que nous voulions évaluer le produit d'une éclaircie du traitement, définie par son intensité ⁽¹⁾ (nombre de tiges ou surface terrière enlevés) et le rapport d'éclaircie r .

On peut admettre que le diamètre D_p prévu par le modèle au moment de cette éclaircie est la moyenne des diamètres avant et après éclaircie :

$$D_p = (D \times D') / 2 \quad (1)$$

D et D' étant les diamètres moyens du peuplement avant et après éclaircie, à déterminer. Mais D et D' sont aussi liés par la valeur du rapport d'éclaircie, ce qui donne une seconde équation :

$$r = (D/D')^2 \quad (2)$$

Le système de deux équations (1) et (2) permet d'obtenir les valeurs de D et D' .

$$D' = 2D_p / (1 + \sqrt{r}) \quad (3)$$

$$D = 2D_p - D' \quad (4)$$

Sachant que :

$$N_e = N - N'$$

$$G_e = G - G'$$

$$G = \frac{\pi}{4} \frac{D^2}{10000} N \quad (\text{si } D \text{ est en cm}) \dots \text{etc.}$$

La connaissance de D et D' permet de calculer facilement la surface terrière enlevée G_e si on connaît le nombre de tiges enlevé N_e (et inversement).

REMARQUE : Quand l'intensité de l'éclaircie est définie par le nombre de tiges enlevé, on peut calculer le diamètre moyen des arbres enlevés en éclaircie à partir de D ou de D' par les formules :

$$D_e = D' \sqrt{\frac{rN - N'}{N - N'}} \quad (5)$$

$$D_e = D \sqrt{\frac{N - N' / r}{N - N'}} \quad (6)$$

Ces relations étant parfaitement indépendantes de l'hypothèse exprimée par la relation (1) ci-dessus, on peut utiliser l'une ou l'autre d'entre elles pour calculer D_e dans le cas où le modèle statique décrit l'évolution du diamètre immédiatement après éclaircie (formule (5)) ou immédiatement avant (formule (6)).

⁽¹⁾ en anglais = thinning intensity

Quand on connaît le diamètre moyen et le nombre des arbres enlevés en éclaircie, il reste à en estimer les volumes total et marchand par application de tarifs de cubage, de peuplement appropriés.

543 Estimation de la production enlevée en éclaircie (modèles dynamiques)

Comme pour les modèles statiques, le traitement sylvicole est défini par l'âge, l'intensité et le rapport d'éclaircie de chaque éclaircie. Si l'intensité est définie par N_e (respectivement G_e), la connaissance de r permet de calculer facilement G_e (respectivement N_e) ⁽¹⁾ : on obtient donc directement les caractéristiques du matériel enlevé en éclaircie et, par soustraction, celles du matériel restant sur pied après l'éclaircie. Cette façon de procéder a déjà été présentée aux § 533 et 534.

Une caractéristique essentielle de nombreux modèles dynamiques est la suivante : l'accroissement du peuplement après une éclaircie dépend de l'intensité de cette éclaircie, dans la mesure où elle détermine la densité du peuplement restant sur pied. Ce n'est pas le cas dans les modèles statiques, d'où le risque d'obtenir, avec ce type de modèle, des résultats incorrects quand on applique un traitement sylvicole qui diffère de ceux sur lesquels on a ajusté les fonctions du modèle.

55 LA MORTALITE

La mortalité peut être considérée comme négligeable dans la plupart des peuplements homogènes gérés de façon intensive, une fois que la plantation est bien installée. Mais il y a des cas où l'on peut observer, indépendamment de toute éclaircie, une nette diminution du nombre de tiges avec le temps, dont un modèle d'accroissement et de production doit rendre compte d'une manière ou d'une autre.

Voyons comment c'est envisageable, selon la nature de la mortalité considérée.

551 Mortalité à l'installation du peuplement ⁽²⁾

On peut définir cette mortalité comme le pourcentage de plants qui n'a pas repris au bout d'un an. Ce pourcentage est facile à évaluer dans des plantations mais, dans le cas de peuplements installés par semis (naturel ou artificiel), il vaut évidemment mieux juger de la réussite de l'installation par le nombre absolu de plants acquis que par un pourcentage de "reprise" ⁽³⁾.

Dans certains cas, on peut définir la reprise par la proportion de la surface qui est effectivement régénérée.

$$(1) \text{ car } r = \frac{G/N}{(G-G_e)/(N-N_e)} = \left(1 - \frac{N_e}{N}\right) \left/ \left(1 - \frac{G_e}{G}\right)\right.$$

(2) en anglais = establishment mortality

(3) en anglais = survival

Les facteurs qui influencent le plus la reprise, et que l'on peut donc envisager d'introduire dans un modèle de prévision de cette reprise, sont :

- la technique de préparation du terrain, et l'intensité de cette préparation,
- les conditions climatiques au moment de l'installation et au cours des mois suivants,
- l'exposition,
- la présence éventuelle et l'importance d'un abri latéral ou vertical,
- la nature du sol en général et de la couche superficielle (10 cm) en particulier.

puis, dans le cas des plantation :

- l'âge des plants,
- l'ensemble des manipulations subies par les plants, de la pépinière à la mise en place,
- la date de la plantation par rapport à la saison de repos de la végétation.

ou, dans le cas de semis (naturels ou artificiels) :

- l'application éventuelle de pesticides ou de répulsifs pour protéger les graines,
- le pré-traitement éventuel des graines pour stimuler ou améliorer la germination.

On se rend bien compte qu'il est virtuellement impossible de déterminer tous ces paramètres et de les inclure dans un modèle quantitatif utilisable. Il vaut mieux construire des modèles simples et robustes où l'on puisse facilement intégrer les pratiques en cours en un lieu donné à un moment donné. Le modèle prendra généralement la forme d'une régression multiple entre la reprise (relative ou absolue) et deux ou trois des paramètres énumérés ci-dessus, codés de manière adéquate.

552 Mortalité liée à la densité

La mortalité liée à la densité ⁽¹⁾ peut être le résultat d'une élimination directe par la concurrence ; elle est plus généralement le fait d'une élimination indirecte : les arbres les moins vigoureux, dont le houppier se trouve dominé, deviennent plus sensibles aux maladies et souffrent plus des attaques d'insectes que les arbres vigoureux.

On peut ignorer ce type de mortalité dans les modèles qui s'appliquent à des peuplements plantés à large espacement et traités à courte révolution, ou soumis à des éclaircies judicieuses. C'est impossible quand le modèle doit s'appliquer à des peuplements non éclaircis ou tardivement éclaircis.

⁽¹⁾ en anglais = density-dependant mortality

Pour en rendre compte, la courbe de REINEKE (voir figure 5.14) peut constituer un modèle de réponse efficace : elle résulte du report sur un graphique à échelle bi-logarithmique de la densité (nombre de tiges vivantes par hectare) d'un peuplement non éclairci en fonction de sa hauteur dominante, et représente la densité maximum supportable par ce peuplement pour une hauteur dominante donnée.

Dans des peuplements normalement éclaircis, il sera difficile d'obtenir les données nécessaires à la construction de cette courbe ; si on ne dispose pas de peuplements non éclaircis, il peut être intéressant, lors d'un boisement, d'installer une dizaine de placettes témoins plantées à 1×1 m ou à 2×2 m sur des stations aussi diverses que possible pour en tirer ces données sur la mortalité naturelle.

La mortalité liée à la densité est implicitement prise en compte dans les modèles de production statiques puisqu'ils ne décrivent que le peuplement sur pied dans les conditions de traitement étudiées : il n'y a donc pas à l'y expliciter.

Dans les modèles d'accroissement dynamiques, on peut simuler la réduction du nombre de tiges due à la mortalité par une éclaircie extrêmement par le bas (voir tableau du § 541) chaque fois que le peuplement dépasse la courbe de REINEKE vers la droite.

553 Mortalité liée aux maladies et aux parasites

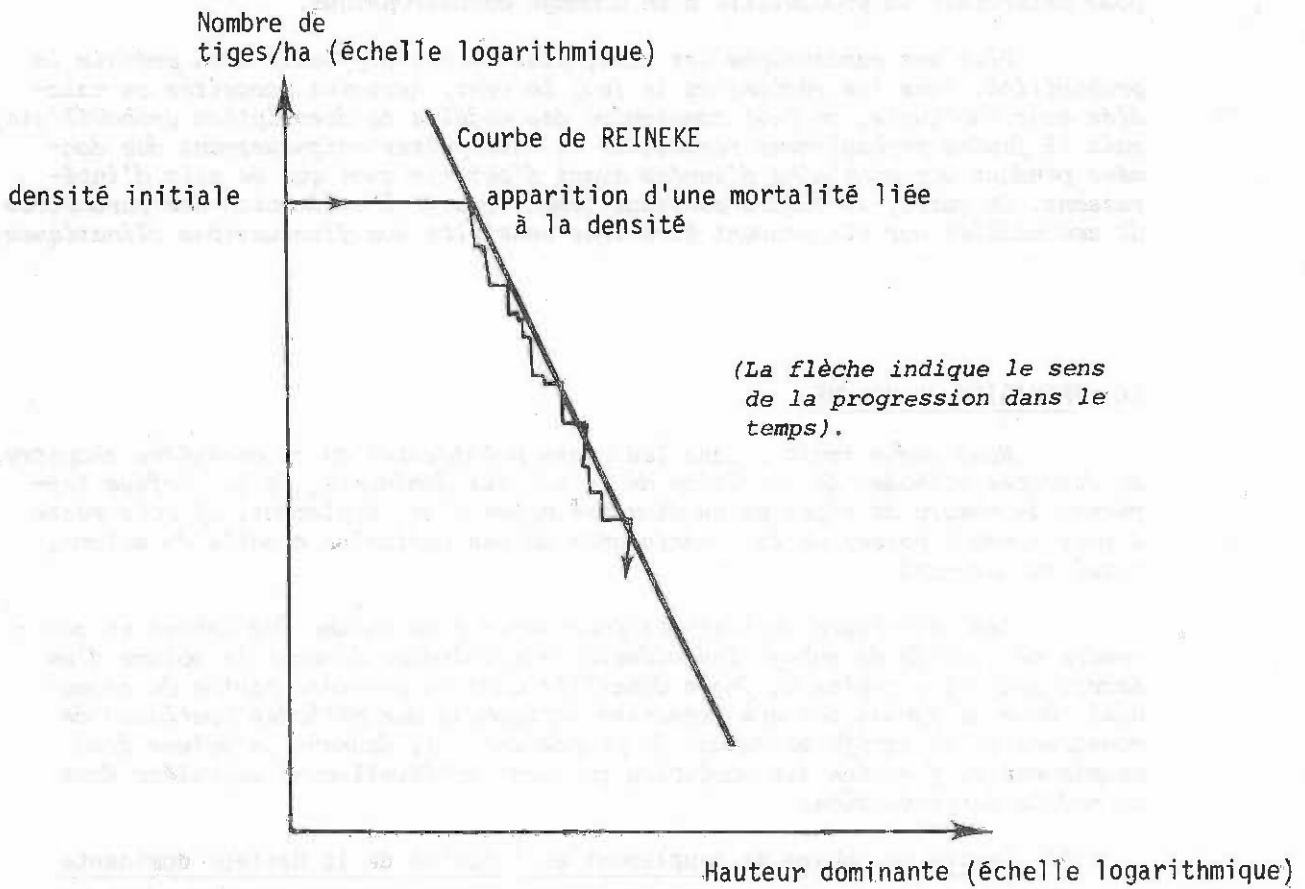
Comme nous l'avons vu au § 552, certains aspects de la mortalité liée aux maladies et aux parasites ⁽¹⁾ peuvent se trouver implicitement inclus dans un modèle simple de mortalité liée à la densité. Mais une maladie ou un parasite peuvent avoir un caractère épidémique, et avoir tendance à se manifester de façon explosive à la suite de séquences climatiques particulières, avec propagation de type contagieux à partir de centres d'infestation. Les techniques de modélisation impliquées sortent du champ de ce manuel ; il s'agit le plus souvent de modèles stochastiques, ne permettant de décrire une explosion qu'en terme de probabilité.

Cette probabilité peut être fonction du climat, de l'état du peuplement, de la proximité d'une autre zone d'épidémie. Pour le gestionnaire, un tel modèle probabiliste peut servir à estimer l'espérance de la valeur de la récolte à une date future, et le coût d'opportunité d'une coupe prématurée, de l'adoption ou de la non adoption de mesures de lutttes contre l'épidémie.

554 Dégâts causés par le vent et le feu

Les dommages causés par le vent, comme ceux dus aux maladies et aux parasites, peuvent être implicitement pris en compte dans la mortalité liée à la densité. Mais ils peuvent avoir un caractère catastrophique - c'est généralement le cas avec le feu - se traduisant par la destruction pure et simple du peuplement. Là encore, la composante catastrophique ne peut être décrite qu'en terme de probabilité d'occurrence, en fonction des conditions climatiques, des caractéristiques de la station et de l'état du peuplement.

(¹) en anglais = disease and pest mortality



Remarque : La courbe de REINEKE est très bien définie pour un peuplement donné, mais la pente et la constante de la droite représentative en échelle bi-logarithmique peuvent varier avec le type de station ; elle peut, dans sa partie inférieure, piquer plus rapidement vers le bas.

Figure 5.14

Evolution dans le temps de la relation entre le nombre de tiges vivantes par hectare et la hauteur dominante dans un peuplement homogène dense non éclairci.

Dans le cas du feu, d'autres facteurs tels que le délai écoulé depuis le dernier feu contrôlé ou la pratique de l'élagage, peuvent être importants pour déterminer la probabilité d'un dommage catastrophique.

Plus une catastrophe est rare, plus il est difficile d'en prévoir la probabilité. Dans les régions où le feu, le vent, certains parasites ou maladies sont habituels, on peut construire des modèles de description probabiliste, mais il faudra probablement recueillir et interpréter soigneusement des données pendant une vingtaine d'années avant d'obtenir quoi que ce soit d'intéressant. En outre, il faudra continuellement revoir l'estimation des paramètres de ces modèles car ils peuvent être très sensibles aux fluctuations climatiques.

56 PREVISION DU VOLUME

Nous avons traité, dans les pages précédentes de ce cinquième chapitre, de diverses méthodes de prévision de la hauteur dominante, de la surface terrière, du nombre de tiges et du diamètre moyen d'un peuplement. Il nous reste à voir comment passer de la connaissance de ces variables à celle du volume, total et marchand.

Les techniques utilisables pour mesurer le volume des arbres et construire des tarifs de cubage individuels (c'est-à-dire donnant le volume d'un arbre) ont été exposées de façon détaillée dans la première partie de ce manuel. Nous ne ferons ici que présenter rapidement des méthodes pratiques de construction de tarifs de cubage de peuplement ⁽¹⁾, donnant le volume d'un peuplement en fonction des variables qui sont habituellement calculées dans un modèle de production.

561 Tarifs de cubage de peuplement en fonction de la hauteur dominante et de la surface terrière

On peut estimer avec précision le volume (total ou jusqu'à une découpe assez petite, par exemple 10 cm de diamètre) d'un peuplement à l'aide d'une équation qui le prédit en fonction de la surface terrière G et de la hauteur dominante H_0 . On peut citer, parmi les modèles les plus courants :

- (1) $V = b_0 + b_1GH_0$ (linéaire)
- (2) $V/G = b_0 + b_1H_0$ (linéaire, pondération en $1/G^2$)
- (3) $\log V = b_0 + b_1 \log G + b_2 \log H_0$ (logarithmique)
- (4) $V/G = b_0 + b_1H_0 + b_2H_0^2$ (quadratique, pondération en $1/G^2$)

et on peut en imaginer bien d'autres.

Quand on ajuste un modèle de prévision du volume aux observations dont on dispose, il est souhaitable de le faire en pondérant les carrés des écarts, parce que l'écart-type résiduel du volume n'est pas constant mais grosso-modo proportionnel au volume lui-même (voir première partie § 353 et annexe A27) : les modèles (2), (3), (4), ci-dessus assurent déjà implicitement une telle pondération, l'écart-type résiduel de V/G ou de $\log V$ étant à peu près constant. Le modèle (2) se prête facilement à un ajustement graphique.

⁽¹⁾ en anglais = stand volume equations

La quantité V/G , qui s'exprime en mètres, est appelée "hauteur de forme".

Avec de tels modèles expliquant le volume total d'un peuplement par sa hauteur dominante et sa surface terrière, on obtient en général un très bon degré d'ajustement, avec un coefficient de corrélation supérieur à 0,99.

562 Utilisation d'un tarif de cubage d'arbres pour estimer le volume d'un peuplement

C'est une pratique assez courante que d'utiliser un tarif de cubage d'arbre pour estimer le volume d'un peuplement en prenant comme valeur des entrées : le diamètre de l'arbre de surface terrière moyenne, et la hauteur moyenne de Lorey du peuplement ; celle-ci est la moyenne des hauteurs des arbres pondérées par leur surface terrière, et peut être estimée par la moyenne pondérée (par la surface terrière) des hauteurs des arbres d'un échantillon systématique. On utilise aussi la hauteur moyenne arithmétique, ou la hauteur dominante...

Dans la plupart des cas, on obtient ainsi une estimation biaisée, parce que les coefficients du tarif individuel ne tiennent pas compte de l'influence des caractéristiques du peuplement sur la distribution en diamètre et sur la forme des arbres. Ce biais, qui va normalement dans le sens d'une sous-estimation du volume du peuplement, est en général assez petit pour que l'on puisse le considérer comme négligeable. Mais on risque une sous-estimation grave du volume du peuplement si le tarif donne le volume jusqu'à une découpe de diamètre fixé et si le diamètre moyen du peuplement est proche de ce diamètre de découpe.

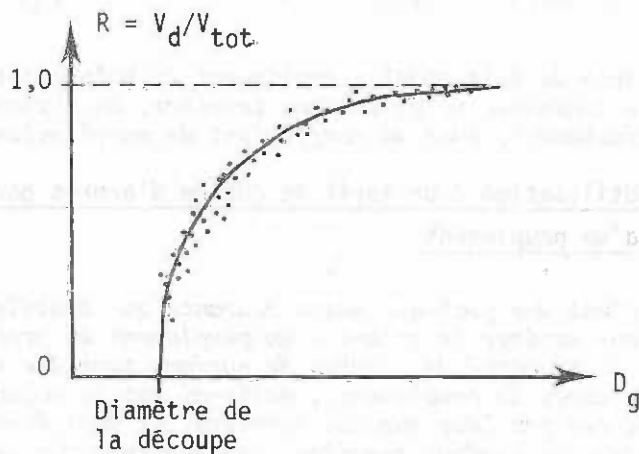
On obtient toujours des résultats plus satisfaisants en utilisant un véritable tarif de cubage de peuplement.

563 Estimation du volume jusqu'à une découpe déterminée ⁽¹⁾

Dans les peuplements homogènes, on a souvent besoin de prévoir la production en volume jusqu'à deux ou trois découpes de diamètres différents. On peut alors procéder de la façon suivante :

- [1] Construire un tarif de cubage de peuplement comme indiqué ci-dessus (voir § 561), donnant le volume total sur écorce (ou sous écorce).
- [2] Calculer, sur chacune des placettes d'échantillonnage mesurées pour la construction du tarif, le rapport entre le volume à la découpe fixée V_d et le volume total V_{tot} .
- [3] Porter sur un graphique les valeurs de ce rapport obtenues, en fonction du diamètre moyen du peuplement ; on obtient un nuage qui a l'allure ci-après :

(¹) en anglais = volume to a top diameter limit



4 On peut y ajuster la courbe :

$$R = 1 - a \exp(b D_g)$$

que l'on peut transformer en :

$$\ln(1-R) = a' + b D_g$$

et ajuster par régression linéaire simple (en notant que $a = \exp(a')$).

Si on procède ainsi pour plusieurs découpes de diamètres différents, on peut essayer d'harmoniser les coefficients obtenus en les exprimant eux-mêmes en fonction du diamètre de la découpe. On peut aussi essayer de construire une extension du modèle ci-dessus en incluant dans les entrées le diamètre de la découpe.

Quand on ajuste le modèle, il faut exclure des données celles où le rapport R est nul.

Quand on utilise le modèle ajusté, il peut donner des valeurs de R négatives (pour les faibles valeurs de D_g) : il faut alors évidemment considérer le volume V_d comme nul.

564 Volume des éclaircies

On peut envisager plusieurs méthodes de calcul du volume des éclaircies :

- On peut utiliser le tarif de cubage du peuplement sur pied, avec en entrée la surface terrière enlevée en éclaircie.
- On peut utiliser un tarif de cubage d'arbre, avec en entrées le diamètre moyen et la hauteur moyenne (estimée à partir d'une régression hauteur/diamètre) des arbres enlevés en éclaircie.

- La méthode la moins biaisée est la construction d'un tarif de cubage "de peuplement" spécial pour les éclaircies, c'est-à-dire d'une régression du volume enlevé en éclaircie sur la surface terrière enlevée et la hauteur dominante du peuplement ; si la régression obtenue n'est pas significativement différente du tarif de cubage du peuplement sur pied, on peut utiliser ce dernier pour cuber les éclaircies.
- Si les volumes enlevés en éclaircie ne sont pas mesurés, il est impossible de construire un tarif spécial pour les éclaircies : on peut alors calculer le volume d'une éclaircie par différence entre les volumes du peuplement avant et après éclaircie, qu'il s'agisse du volume total ou du volume marchand (à une découpe fixée).

57. TECHNIQUES AVANCEES DE PREVISION DE L'ACCROISSEMENT ET DE LA PRODUCTION

Il y a deux types de modèles, applicables aux peuplements homogènes, dont nous n'avons pas encore parlé ; plus sophistiqués, ils sont déjà utilisés dans un certain nombre d'organismes forestiers : ce sont les modèles "par catégories de grosseur" ⁽¹⁾ et les modèles de "croissance individuelle d'un arbre soumis à la concurrence individuelle de ses voisins". ⁽²⁾.

571 Modèles par catégories de grosseur

Dans un modèle par catégories de grosseur construit pour un peuplement homogène, le matériel sur pied est décrit à tout moment par la distribution des arbres en catégories de grosseur (diamètre à 1,30 m). L'évolution de l'effectif de chaque catégorie est calculée séparément, généralement en fonction de la fertilité, de l'âge, de la densité du peuplement et du rapport entre le diamètre représentatif de la catégorie et le diamètre moyen (ou dominant) du peuplement.

Les éclaircies et la mortalité sont également décrites en termes de distribution, chaque éclaircie ou manifestation de la mortalité modifiant en conséquence la distribution du matériel sur pied.

L'avantage de ce type de modèles dans les peuplements homogènes est de fournir plus de détails sur la ventilation de la production par catégorie de grosseur, surtout si on utilise une fonction décrivant la forme des arbres pour les calculs de volume. L'inconvénient en réside dans la plus grande complexité de l'analyse des données concernant la croissance des peuplements étudiés, nécessitant l'introduction de variables supplémentaires pour décrire les distributions en diamètre. Leur utilisation exige d'avoir accès à un ordinateur.

Par rapport aux modèles de peuplement déjà présentés dans les paragraphes précédents, ces modèles fournissent donc une information plus détaillée sur le peuplement, mais pas forcément une prévision plus précise de ses caractéristiques globales (surface terrière, diamètre moyen, ...).

⁽¹⁾ en anglais = size class models

⁽²⁾ en anglais = tree position models

572 Modèles de croissance individuelle d'un arbre soumis à la concurrence individuelle de ses voisins

Ce sont des modèles dans lesquels les relations spatiales entre tous les arbres d'une placette simulée sont réellement prises en compte. La concurrence entre arbres dépend en effet des tailles et positions relatives des arbres voisins. L'étude de la compétition à l'intérieur des peuplements permet de définir un indice de concurrence, calculable pour chaque arbre en fonction de ses caractéristiques et de celles de ses voisins : cet indice, appliqué comme coefficient de réduction à l'accroissement qu'aurait l'arbre en l'absence de toute concurrence (accroissement dépendant au moins de l'âge et de la fertilité), fournit son accroissement réel dans les conditions de concurrence où il se trouve.

Le fonctionnement de ces modèles demande une masse énorme de calculs et peut mobiliser toute la capacité d'un gros ordinateur. Leur complexité est telle qu'il n'est pas intéressant de les utiliser pour modéliser la croissance de peuplements homogènes : ils ne fournissent pas beaucoup plus d'information utile sur la croissance du peuplement qu'un modèle par catégories de grosseur qui demande cent fois moins de calculs. En revanche, ils peuvent être très utiles dans les recherches sur la physiologie des arbres, ainsi que dans la prévision de l'accroissement et de la production de peuplements hétérogènes.

Les lecteurs qui désirent un supplément d'information sur ces deux types de modèle peuvent se reporter à la bibliographie commentée, éditée par le Commonwealth Agricultural Bureau sous le titre "Computerized methods in forest planning and forecasting" (voir bibliographie en annexe D).

6 ANALYSE DES DONNEES D'ACCROISSEMENT ET DE PRODUCTION (PEUPELEMENTS HETEROGENES)

Au sens où il est utilisé dans ce manuel, le terme de "peuplement hétérogène" s'applique quand le peuplement qui constitue l'unité de mesure de base comprend plusieurs classes d'âge. Généralement, il est en outre constitué de plusieurs espèces :

- soit écologiquement semblables,
- soit appartenant à différents groupes écologiques, dont chacun tend à dominer dans une strate du couvert ou sur certaines microstations ou dans l'une des phases suivant la formation d'une trouée ou une exploitation.

La distinction entre peuplement hétérogène et peuplement homogène n'est pas absolue. Quand un peuplement hétérogène n'est constitué que de deux classes d'âge -un étage supérieur dominant un sous-étage de régénération très jeune-, l'exploitation se traduisant alors en gros par une coupe rase, on peut lui appliquer les techniques de prévision de la production des peuplements homogènes. Inversement, un peuplement équienne composé de plusieurs essences devra être analysé par des techniques de peuplements hétérogènes, si l'on veut pouvoir rendre compte dans un modèle unique de différents mélanges se distinguant par la proportion de chaque essence.

Les peuplements hétérogènes posent généralement des problèmes de mesure et d'échantillonnage beaucoup plus ardues que les peuplements homogènes. A l'exploitation, la valeur de l'hectare de peuplement hétérogène (en forêt plus ou moins "naturelle") sera généralement plus faible que celle de l'hectare de peuplement homogène arrivé à maturité : il sera difficile d'y justifier une intensité d'échantillonnage équivalente, alors que la dispersion des essences de valeur et l'hétérogénéité du peuplement demanderaient un échantillonnage beaucoup plus important pour arriver à une prévision aussi précise.

Les problèmes posés par les difficultés d'accès, la densité du sous-étage, l'existence de contreforts et la quasi-impossibilité dans bien des cas de mesurer la hauteur totale des arbres rendent bien plus coûteuse l'obtention d'une précision donnée sur les mesures.

Ceci contribue encore à réduire l'échantillonnage réalisable dans le cadre d'un budget fixé.

Tout ceci explique que la plupart des études de production en forêt hétérogène aient été faites à partir de données en nombre relativement faible et les modèles qui en résultent ont naturellement une précision bien inférieure à celle que l'on exige dans la gestion de peuplements homogènes. Mais il y a des exceptions, comme en témoigne l'étude de production relatée par REVILLA ⁽¹⁾ aux Philippines : il y cite des résultats sur la forêt à diptérocarpées obtenus à partir de plus de 240 placettes permanentes d'échantillonnage.

(¹) A.R. REVILLA, 1979, "Yield prediction in cut-over Dipterocarp stands in the Philippines", paper to seminar on management of Dipterocarp forests, Metro Manila 20 pages.

Vu la diversité de la composition en essences, de la structure de la flore, des situations écologiques et des pratiques sylvicoles que l'on rencontre en forêt hétérogène, les stratégies de modélisation applicables y sont encore plus nombreuses que dans les peuplements homogènes. La plupart des méthodes qui ont fait l'objet d'une publication n'ont pas été testées, si ce n'est sur la base d'un très petit nombre de placettes. Les techniques que l'on a choisi de présenter ici sont probablement les plus robustes ; on peut les classer en trois groupes :

- les modèles de production statiques applicables aux types de peuplements hétérogènes les plus simples,
- les modèles utilisant des matrices de transition,
- les modèles de croissance individuelle d'un arbre soumis à la concurrence du peuplement.

Un quatrième type de modèles, les modèles de croissance individuelle d'un arbre soumis à la concurrence individuelle de ses voisins, pourrait constituer dans l'avenir l'outil le plus efficace de prévision de la production des peuplements hétérogènes. Leur utilisation demande malheureusement des ordinateurs plus puissants que ceux dont on dispose généralement aujourd'hui, et ils ne seront concrètement utilisables qu'après l'aboutissement de gros efforts de recherche sur les techniques d'échantillonnage appropriées.

Ce chapitre ne veut pas donner d'instructions précises sur la construction de modèles de croissance et la prévision de la production dans les peuplements hétérogènes, mais plutôt quelques exemples illustrant diverses façons d'aborder ces problèmes. Vu la masse de travaux en cours dans ce domaine, on peut espérer que des méthodes valables seront bientôt mises au point ...

61 CLASSEMENT DE LA STATION (détermination de la fertilité)

Les peuplements hétérogènes étant généralement constitués de nombreuses essences différentes, le concept d'indice de fertilité doit être associé à une essence (modèles de croissance individuelle) ou à un groupe d'essences (modèles de peuplement).

De nombreuses techniques de détermination de la fertilité ont été proposées, s'appuyant toutes sur les mêmes principes ; les indicateurs de la fertilité sont :

- soit des fonctions de paramètres dendrométriques (comme la hauteur dominante et l'âge dans les courbes de fertilité des peuplements homogènes),
- soit des facteurs du milieu comme l'altitude, la nature du sol, la pluviométrie, ...etc, ou des indicateurs indirects du milieu comme la présence d'espèces caractéristiques ou l'appartenance à une strate géographique définie.

Les indicateurs de la fertilité appartenant à la première catégorie sont fondamentalement des écarts observés entre un peuplement réel et un modèle dans lequel la fertilité n'est pas incluse : le système de courbes de fertilité d'un type de peuplements homogènes ne fait que concrétiser les écarts que l'on

peut observer entre la croissance en hauteur d'un peuplement donné et une courbe "moyenne" de croissance en hauteur toutes fertilités confondues.

Ainsi, si l'on conçoit un modèle applicable à un type de peuplements hétérogènes et prédisant la valeur d'une certaine variable Y sans tenir compte de la fertilité, les valeurs \hat{y}_i prédites par le modèle sur un certain nombre de placettes seront plus ou moins différentes des valeurs y_i observées sur ces placettes : les écarts résiduels $\hat{y}_i - y_i$ ainsi obtenus peuvent être codés sur une échelle du genre 1,2,3,4,5,...,etc, avec une valeur médiane représentant l'écart nul ($\hat{y} = y$), et l'on disposera ainsi d'un système de mesure de la classe de fertilité applicable aux placettes observées.

On se sert d'un tel système de classes de fertilité en déterminant les écarts (indicateurs de la fertilité) à un instant t et en les utilisant, comme entrée supplémentaire du modèle, pour faire une prévision à l'instant $t + 1$. C'est bien ce que l'on fait avec un système classique de courbes de fertilité appliqué à une plantation : on détermine la classe de fertilité à partir d'une mesure hauteur-âge à un instant donné, puis on l'introduit dans le modèle pour obtenir une prévision de la hauteur à un instant ultérieur bien plus précise que celle que fournirait la courbe moyenne de croissance en hauteur.

Une fois que l'on a déterminé, sur un certain nombre de placettes, les écarts résiduels à un modèle ignorant la fertilité, on peut les corrélérer à des variables du milieu pour mettre en évidence celles qui expliquent le mieux la variation résiduelle par rapport au modèle (indicatrice de la fertilité). L'utilisation de variables conditionnelles prenant les valeurs 0 ou 1 (voir annexe A) permet d'intégrer dans le modèle des variables purement qualitatives comme la présence d'une espèce caractéristique, ou l'appartenance à un type de sol ... etc.

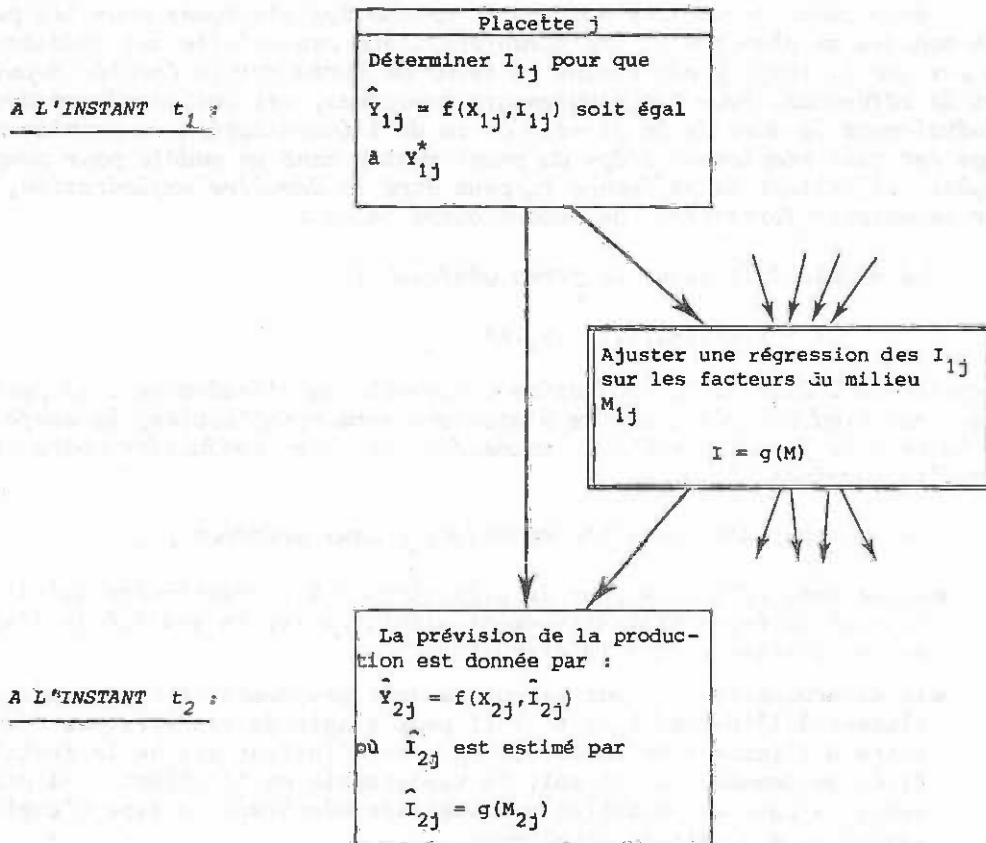
Il est plus difficile d'utiliser des variables dendrométriques comme indicateurs de la fertilité dans les peuplements hétérogènes que dans les peuplements homogènes, et ce n'est possible en fait qu'avec les modèles de production statiques. En revanche, les variables du milieu peuvent être utilisées comme indicateurs de la fertilité dans n'importe quel type de modèle : c'est donc à elles qu'on fera presque toujours appel dans les peuplements hétérogènes, alors que leur présence dans les modèles de peuplement homogène, où la croissance en hauteur est la base de la détermination de la fertilité, n'est qu'un raffinement.

D'après tout ce qui vient d'être dit, on comprendra que les termes conventionnels "indice de fertilité" et "classe de fertilité" ne sont pas les plus appropriés pour décrire les concepts en jeu. Il s'agit en fait de modèles à auto-ajustement qui utilisent une évaluation de leur erreur à un instant donné pour réduire l'incertitude d'une prévision ultérieure.

Mais il n'y a rien d'intrinsèquement uni-dimensionnel dans l'effet du milieu sur la croissance de la forêt. Par ailleurs, une partie de l'écart corrigé par le processus d'auto-ajustement peut très bien être dû à tout autre chose que le milieu, par exemple à des facteurs historiques (sylviculture passée), à la densité ou aux caractéristiques génétiques du peuplement, si tout cela n'est pas pris en compte dans le modèle. Cette méthode ne permet donc pas forcément de décrire "toute la fertilité et rien que la fertilité".

Ceci peut être résumé symboliquement de la façon suivante :

Etant donné le modèle d'accroissement
ou de production $Y = f(X, I)$:



Symboles

- M : ensemble des variables du milieu utilisées dans l'ajustement $I = g(M)$
(M_{ij} = valeurs de cet ensemble à l'instant t_i sur la placette j).
- f () : modèle mathématique utilisé pour prévoir l'accroissement ou la production.
- g () : fonction mathématique utilisée pour la régression de I sur les variables du milieu.
- I_j : indice de fertilité, déterminé sur chaque placette à l'instant t_1 par annulation de l'écart entre productions observée et prévue par le modèle (I_{1j} = sur la placette j).
- \hat{I} : indice de fertilité estimé en fonction des variables du milieu par la régression $I = g(M)$ (\hat{I}_{2j} : à l'instant t_2 sur la placette j)
- X : ensemble des variables explicatives du modèle de production (X_{ij} = valeurs de cet ensemble à l'instant t_i sur la placette j).
- \hat{Y} : production estimée par le modèle (\hat{Y}_{ij} = à l'instant t_i sur la placette j).
- Y_1^* : production observée à l'instant t_1 (Y_{1j}^* : sur la placette j).

62 MODELES DE PRODUCTION STATIQUES EN FORET HETEROGENE

621 Principes généraux

Nous avons traité des modèles de production statiques pour les peuplements homogènes au chapitre 5. Une caractéristique essentielle des modèles statiques est que le temps y est inclus en terme de durée totale écoulée depuis un instant de référence. Pour les peuplements homogènes, cet instant de référence est généralement la date de la plantation ou de l'installation du peuplement, et le temps est tout simplement l'âge du peuplement ; dans un modèle pour peuplements hétérogènes, l'instant de référence t_0 peut être la dernière exploitation, le dernier inventaire forestier, ou encore autre chose...

Le modèle doit avoir la forme générale :

$$Y = f(x_1, x_2, x_3, \dots, x_n, t)$$

où Y représente l'état de la production à laquelle on s'intéresse ; ce peut être le volume sur pied de bois d'oeuvre d'essences commercialisables, la surface terrière, le bois de feu disponible en tonnes/ha, ou toute production autre que le bois (fruits, graines, écorce...).

Les x_i sont des variables répondant à deux critères :

- elles sont efficaces pour la prévision de Y , c'est-à-dire qu'elles améliorent de façon statistiquement significative la qualité de l'ajustement de la fonction $f()$ à la production.
- la détermination de leur valeur ne doit pas demander une mesure du peuplement à l'instant $t_0 + t$: il peut s'agir de statistiques d'inventaire à l'instant de référence t_0 , ou d'indicateurs de la fertilité tirés de données sur le sol, la topographie ou le climat ; il peut également s'agir de variables qualitatives décrivant le type d'exploitation auquel on a soumis le peuplement.

La variable t est le temps écoulé depuis l'instant de référence ; elle doit figurer dans la fonction si l'on veut que celle-ci constitue un instrument de planification utile.

La forme de la fonction peut varier considérablement selon la nature de la production estimée, les données utilisées et la puissance des outils d'analyse dont on dispose.

622 Analyse des données

Les données peuvent provenir de placettes temporaires ou permanentes, ou d'expériences. La méthode d'analyse des données la plus simple que l'on puisse envisager est la régression linéaire multiple, avec analyse graphique des données et des résidus : il est donc indispensable de disposer d'un ordinateur, même s'il n'a pas besoin d'être très puissant.

Les données provenant des placettes doivent être résumées avant de s'attaquer à la détermination de l'équation de production. Le résumé d'une placette contiendra les informations suivantes pour chaque mesure :

- 1] Identification de la placette ; pour les expériences, elle précisera le numéro de bloc, le numéro du traitement et le numéro de répétition.
- 2] Les "productions" sur pied ; on peut ne s'intéresser qu'à un type de production (par exemple : le volume marchand toutes essences), mais généralement on veut la détailler par groupes d'essences et catégories de grossueur et de qualité : on arrive alors facilement à une dizaine de variables "production".
- 3] Les variables explicatives possibles, parmi lesquelles :
 - . la surface terrière/ha par groupes d'essences à l'instant t_0 ,
 - . la proportion en nombre de tiges de chaque groupe d'essences à l'instant t_0 ,
 - . l'intensité de l'exploitation à l'instant t_0 ,
 - . un code pour le traitement sylvicole,
 - . un code pour le type de peuplement,
 - . des facteurs du milieu (type de sol, altitude, pluviométrie, etc).

Ces "résumés de placettes" sont stockés sous une forme accessible à l'ordinateur (voir chapitre 4) pour les analyses qui suivront.

Pour effectuer ces analyses, on peut utiliser diverses méthodes telles que l'analyse en composantes principales, la régression progressive, la régression "combinatoire" ⁽¹⁾, aboutissant à une sélection plus ou moins automatique des variables explicatives qui expliquent "le mieux" l'état des productions. Toutefois, une sélection a priori de variables explicatives donnant une équation relativement simple est généralement préférable, quand on l'associe à une analyse graphique serrée des résidus ; le bon sens peut aider au choix de la forme de la fonction en la reliant aux connaissances que l'on a sur la réalité biologique étudiée.

623 Choix de la forme de l'équation

On a présenté au chapitre 5 différentes formes d'équations utilisables pour modéliser des processus de croissance asymptotique, et dans l'annexe A (figure A22) d'autres formes non asymptotiques. Lorsqu'on recherche une courbe ou une fonction modélisant l'évolution de la ressource sur pied en forêt naturelle, il faut commencer par représenter graphiquement les données : état de la production Y en fonction du temps t écoulé depuis l'instant de référence t_0 (exploitation ou inventaire). Les données doivent être différenciées par catégories de surface terrière à l'instant t_0 , et l'on peut faire des graphiques distincts pour les différents types de station, ou de forêt (se distinguant par leurs compositions en essences).

⁽¹⁾ en anglais = combinatorial regression ; la méthode consiste, à partir d'un ensemble de n variables explicatives "proposées", à faire les régressions sur tous les sous-ensembles possibles de 1, de 2, ... de n variables explicatives, et à retenir celle qui présente le plus fort coefficient de corrélation multiple. On est ainsi conduit à faire $2^n - 1$ régressions, ce qui devient vite prohibitif quand n augmente.

De tels graphiques peuvent aider au choix d'une équation appropriée parmi les diverses formes décrites dans ce manuel ; on peut alors l'ajuster directement aux données par régression multiple ou par régression sériée. On peut aussi tracer une courbe à la main dans le nuage des données, et lui ajuster ensuite une équation si besoin est.

Mais bien souvent, la forêt est un mélange complexe d'essences et de types écologiques variés, et l'exploitation est très sélective donc légère : il se peut alors qu'il n'y ait aucune relation discernable entre l'état de la production et la surface terrière initiale, la classe de fertilité, le temps. Dans ce cas, les modèles de production statiques ne constituent pas une stratégie efficace de prévision de l'accroissement.

624 Problèmes posés par les modèles de production statiques dans les peuplements hétérogènes

On pourrait croire que les modèles de production statiques, parce qu'ils sont relativement simples à construire et à utiliser, constituent une sorte de panacée pour résoudre les problèmes de prévision de la production dans les peuplements hétérogènes. C'est malheureusement loin d'être le cas. Les problèmes qui se posent bien que plus aigus, sont fondamentalement de même nature que ceux que soulèvent les modèles de production statiques pour peuplements homogènes (voir § 525). Ils sont essentiellement, outre l'absence possible déjà signalée de toute relation significative, de deux types :

- Un modèle de production donné contient implicitement une évolution historique correspondant à celle, si elle n'est pas trop hétérogène, des données utilisées pour sa construction. On ne peut donc pas l'appliquer avec confiance à des peuplements qui ont subi une histoire différente.
- Il se pose un problème de compatibilité quand le modèle intéresse plusieurs productions. Supposons par exemple que l'on prédise trois volumes différents : V_c pour les essences commercialisables, V_p pour les essences partiellement commercialisables et V_i pour les essences inutilisées ; la somme de ces trois volumes devrait en principe représenter le volume de tout le peuplement. Or, si l'on compare $(V_c + V_p + V_i)$ au volume toutes essences, tiré de données réelles ou d'une quatrième fonction ajustée directement à ce volume, on constatera souvent un biais considérable.

Aucun de ces deux problèmes n'est en principe insurmontable. Si l'on dispose en quantité suffisante de données provenant de placettes permanentes et d'expériences à long terme, on peut intégrer au modèle des variables qualitatives représentant différents types d'histoire sylvicole. On peut résoudre les problèmes de compatibilité en ajustant des régressions sous contrainte les obligeant à vérifier certaine condition.

Les modèles de production statiques souffrent en outre de ne pas pouvoir utiliser des données provenant de sources variées (par exemple : expériences de types divers, placettes permanentes d'échantillonnage et placettes de mesure des accroissements individuels). C'est là une limitation sérieuse.

625 Conclusions sur les modèles de production statiques dans les peuplements hétérogènes

- Les modèles de production statiques sont relativement simples à construire si l'on possède les données appropriées.
- Dans de nombreux types de peuplements hétérogènes, il n'existe pas de relation nette entre la production et le temps, l'intensité du traitement, la fertilité : il est alors impossible d'utiliser des modèles statiques.
- Lorsqu'on utilise un modèle statique, on ne doit pas oublier qu'il ne peut rendre compte correctement que des histoires sylvicoles couvertes par les données qui ont servi à sa construction.
- Quand on veut prédire plusieurs catégories de production, il faut ajuster des régressions sous contrainte, ou alors accepter que la somme de ces estimations soit biaisée par rapport au total réel.

63. MODELES UTILISANT DES MATRICES DE TRANSITION

Pouvant être construits sur des données de qualité médiocre mais en grand nombre tirées de mesures successives, les modèles de transition ⁽¹⁾ servent à prédire l'évolution à court terme du peuplement sur pied.

Les données peuvent être des données d'inventaire forestier continu tirées de placettes permanentes ou même de placettes temporaires successives situées à peu près au même endroit, sur lesquelles les arbres ont été inventoriés par catégories de diamètre et groupes d'essences. Elles peuvent même contenir un nombre non négligeable d'erreurs de mesure sans que le modèle en soit trop affecté.

Dans le contexte de la forêt hétérogène, le "court terme" peut couvrir jusqu'à 40 ou 50 ans, ou un cycle d'exploitation.

La méthode est objective, mais pas forcément très précise. Son principal avantage est de s'accommoder d'une analyse automatique des données, qui peuvent être des données d'inventaire absolument conventionnelles. Son principal inconvénient est d'exiger généralement, si l'on veut un modèle qui ne soit pas trop grossier, une grosse masse de données.

631 Définition

Ce type de modélisation suppose que l'état du système étudié est représenté par une série de n variables appelée vecteur d'état ⁽²⁾. En forêt, on peut imaginer le vecteur d'état :

$$X_m = (x_{m1}, x_{m2}, \dots, x_{mi}, \dots, x_{mn})$$

où chaque x_{mi} représente le nombre d'arbres de la catégorie de diamètre i , le premier indice m indiquant l'instant où l'état du système est décrit.

⁽¹⁾ en anglais = transition models

⁽²⁾ en anglais = state vector

Les x_i peuvent représenter tout autre chose que des nombres par catégorie de grosseur, par exemple les biomasses des différents niveaux trophiques d'un écosystème ou les nombres d'individus ayant atteint différents stades de développement dans une population animale. Mais dans les modèles d'accroissement en forêt hétérogène, il s'agira généralement d'une description par catégorie de grosseur.

La matrice de transition ⁽¹⁾ T est un tableau de dimension $n \times n$:

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & \dots & t_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ t_{n1} & t_{n2} & \dots & \dots & t_{nn} \end{bmatrix}$$

Chaque élément t_{ij} représente la proportion de x_{mi} qui devient $x_{m+1,j}$ dans l'intervalle de temps $[m, m+1]$: la valeur de $x_{m+1,j}$ sera la somme des x_{mi} multipliés par les t_{ij}

$$x_{m+1,j} = \sum_{i=1}^n t_{ij} x_{mi} \quad (1)$$

On appelle t_{ij} : probabilité de transition ⁽²⁾ de la catégorie i à la catégorie j dans l'intervalle $[m, m+1]$. En utilisant l'équation (1) pour chaque valeur de j (de 1 à n), on obtient à partir de X_m les composantes du nouveau vecteur d'état X_{m+1} . En notation matricielle, ceci s'écrit simplement :

$$X_{m+1} = X_m T$$

ILLUSTRONS CECI PAR UN EXEMPLE :

Supposons que les composantes du vecteur d'état soient les nombres de tiges des catégories de diamètre suivantes :

<20, 20-40, 40-60, 60-80, ≥ 80

et que la matrice de transition soit :

(1) en anglais = transition matrix
 (2) en anglais = transition probability

		Catégories de diamètre en fin de période					Total
		< 20	20-40	40-60	60-80	≥ 80	
Catégories de diamètre en début de période.	< 20	0,60	0,39	0,01	0	0	1
	20-40	0	0,68	0,30	0,02	0	1
	40-60	0	0	0,75	0,21	0,04	1
	60-80	0	0	0	0,90	0,10	1
	≥ 80	0	0	0	0	1,00	1

La présence de zéros en-dessous de la diagonale est liée au fait que les arbres ne peuvent pas quitter une catégorie de diamètre pour une catégorie inférieure, car ils ne maigrissent pas. La somme des probabilités de transition de chaque ligne doit être égale à 1 si le vecteur d'état peut vraiment définir tous les états possibles du système.

Supposons maintenant que la distribution des diamètres au début de la période soit la suivante :

Catégorie de diamètre	< 20	20-40	40-60	60-80	≥ 80
Nombre de tiges/ha	1037	219	88	22	7

Il suffit de construire le tableau ci-dessous, où la colonne de droite comprend l'effectif de chaque catégorie à l'instant m ; les autres chiffres du tableau sont obtenus, pour chaque ligne, en multipliant cet effectif par la probabilité de transition correspondante :

		Catégories de diamètre en fin de période					Total
		< 20	20-40	40-60	60-80	≥ 80	
Catégories de diamètre en début de période.	< 20	623	404	10			1037
	20-40		149	66	4		219
	40-60			66	19	3	88
	60-80				20	2	22
	≥ 80					7	7
Total		623	553	142	43	12	

Le total de chaque colonne (ligne inférieure) donne l'effectif de chaque catégorie à l'instant $m+1$.

Cet exemple ignore le passage à la futaie ⁽¹⁾, la mortalité et l'exploitation, dont il sera traité dans les paragraphes suivants.

(¹) en anglais = ingrowth

632 Méthodes de construction des matrices de transition

Nous décrivons ici deux méthodes de construction de la matrice de transition sur l'intervalle $[m, m+1]$ d'un peuplement connu par deux mesures aux instants m et $m+1$. On emploie la première méthode quand les arbres du peuplement ont été suivis individuellement, et la seconde quand le peuplement est décrit par catégories de grosseur.

632.1 A partir de données où les arbres sont suivis individuellement

Nous allons décrire la construction de la matrice de transition, sur l'intervalle de temps séparant deux mesures, d'un peuplement dont on a suivi individuellement les arbres de la première à la seconde mesure. On connaît dans ce cas, entre autres choses :

- La liste des arbres apparus dans le peuplement, c'est-à-dire ayant atteint entre les deux mesures la dimension minimum d'inventaire (arbres "passés à la futaie"), avec leur diamètre au moment de la seconde mesure ;
- la liste des arbres disparus parce qu'ils sont morts entre les deux mesures, avec leur diamètre au moment de la première mesure ;
- la liste des arbres disparus parce qu'ils ont été exploités entre les deux mesures, avec leur diamètre au moment de la première mesure ;
- et enfin la liste des arbres présents lors des deux mesures, avec leur diamètre au moment de chacune des deux.

La matrice de transition a la structure ci-dessous où R est une catégorie "régénération" contenant les arbres trop petits pour être inventoriés, a, b, c, ..., n les catégories de diamètre successives (la dernière n'étant pas fermée à droite), M une catégorie "mortalité" qui reçoit les arbres morts et E une catégorie "exploitation" qui reçoit les arbres exploités.

		Deuxième mesure :									Total
		R	a	b	c	d	n	M	E	
Première mesure :	R	x	x	x	x	x				
	a	0	x	x	x	x	x	x	x	
	b	0	0	x	x	x	x	x	x	
	c	0	0	0	x	x	x	x	x	
	d	0	0	0	0	x	x	x	x	

	n	0	0	0	0	0	x	x	x	x
M	0	0	0	0	0	0		0		
E	0	0	0	0	0	0	0			

Elle se construit de la façon suivante :

- 1 La catégorie initiale (première mesure) et la catégorie finale (deuxième mesure) de chaque arbre sont connues : on pointe chaque arbre dans la case située à l'intersection ligne \times colonne correspondante, ainsi que dans la colonne "total" à droite.
Seules les cases marquées d'une croix pourront se trouver remplies, à cause des contraintes biologiques et de la nature des données. De nombreuses cases contiennent un zéro par définition ; quelques-unes ne peuvent pas être remplies, ce qui pose des problèmes pour l'interprétation de la première ligne comme nous allons le voir plus loin.
- 2 Les probabilités de transition t_{ij} s'obtiennent en divisant le nombre d'arbres pointés dans chaque case ij par le nombre total d'arbres de la ligne (colonne "total"). Il n'y a pas de problème pour les lignes correspondant aux catégories de diamètre, ni pour les deux dernières lignes où l'on peut donner la valeur 1 aux probabilités de transition t_{MM} et t_{EE} (un arbre mort, ou exploité, lors de la première mesure ne ressuscitera pas).
- 3 L'effectif de la catégorie "régénération" est généralement inconnu, comme ses transitions vers les catégories M et E. On peut imaginer plusieurs façons de réagir à cette situation, entre autres :
 - (i) essayer d'estimer l'effectif de la "régénération" à la première mesure et diviser le contenu des cases de la ligne par cet effectif : on aura bien ainsi une estimation des probabilités de transition de R vers les catégories de diamètre a, b, c, ... etc.
 - (ii) considérer la catégorie R non plus comme la "régénération" mais comme une catégorie non dimensionnelle, donc de nature différente des autres, contenant tous les arbres passés à la futaie pendant l'intervalle entre les deux mesures ; le nombre total de ces arbres est connu, mais les probabilités t_{Ra} , t_{Rb} , ..., ainsi obtenues ne représentent plus à proprement parler des transitions, mais une simple ventilation du passage à la futaie entre les diverses catégories de diamètre.
 - (iii) considérer que les petits arbres sont "issus" des gros, c'est-à-dire supprimer la ligne et la colonne R et pointer ces petits arbres sur la ligne n sans les pointer dans la colonne Total ; on aboutit alors souvent à des t_{na} , t_{nb} , ... qui peuvent être supérieurs à 1 et on ne peut plus décemment parler pour cette nouvelle ligne n de "probabilité de transition", mais plutôt de "coefficient de régénération" ou autre terme analogue...

632.2 A partir de données décrivant le peuplement par catégories de grosseur

Le principe de construction reste le même, mais le fait de disposer d'une information de moindre qualité oblige à faire des hypothèses assez fortes, qui ont d'autant moins de chances d'être vérifiées que l'intervalle séparant les deux mesures est plus long et les catégories de grosseur plus fines :

- tous les arbres passés à la futaie sont restés dans la plus petite catégorie ;
- les arbres ne peuvent que rester dans la même catégorie ou passer dans la catégorie immédiatement supérieure ;
- le nombre d'arbres exploités dans chaque catégorie est connu par ailleurs, ou déduit des règles de culture.

La matrice de transition aura alors la structure suivante :

		Deuxième mesure :						A
		R	a	b	n	E	
Première mesure :	R	•	•	0	0	0	?
	a	0	•	0	0	0	
	b	0	0	•	$T_{i-1,i}$	0	
	⋮	0	0	0	T_{ii}	$T_{i,i+1}$	E_i
	⋮	0	0	0	0	0	A_i
	n	0	0	0	0	0	
	E	0	0	0	0	0	[1]
	B				B_i			

Seules les cases de la matrice de transition situées sur la diagonale, immédiatement à droite de la diagonale et dans la colonne E peuvent contenir autre chose que zéro. On indique dans la colonne A les nombres de tiges par catégories de diamètre à la première mesure, dans la ligne B les nombres de tiges par catégories de diamètre à la deuxième mesure et dans la colonne E les nombres de tiges par catégories de diamètre exploités entre les deux mesures. On détermine alors de proche en proche, en partant de la plus grosse catégorie, les transitions en nombre de tiges $T_{i-1,i}$ et T_{ii} à l'aide des formules suivantes :

$$T_{ii} = A_i - E_i - T_{i,i+1} \quad (2)$$

$$T_{i-1,i} = B_i - T_{ii} \quad (3)$$

en démarrant avec $T_{n,n+1} = 0$ dans la formule (2), et en s'arrêtant à la deuxième colonne (première catégorie de diamètre). Les problèmes posés par la première et la dernière ligne ont déjà été discutés au paragraphe précédent.

Il reste alors, pour obtenir les probabilités de transition, à diviser par A_i les T_{ii} , $T_{i,i+1}$ et E_i . Les probabilités obtenues représentent :

t_{ii} : la proportion des arbres de la catégorie i qui y restent,

$t_{i,i+1}$: la proportion des arbres de la catégorie i qui passent dans la catégorie immédiatement supérieure,

t_{iE} : la proportion des arbres de la catégorie i qui sont exploités,
tout ceci pendant l'intervalle $m, m+1$ séparant les deux mesures.

On peut très facilement ajouter à la matrice une catégorie "mortalité" M , à condition de connaître le nombre d'arbres de chaque catégorie de diamètre qui sont morts dans l'intervalle $[m, m+1]$. Elle se traite exactement comme la catégorie "exploitation" : il suffit de soustraire M_i au second membre de l'équation (2).

Cette méthode peut être utilisée pour déterminer la matrice de transition d'une forêt sur l'intervalle séparant deux inventaires successifs.

Il faut bien noter maintenant que, quelle que soit la méthode utilisée pour construire la matrice de transition d'un peuplement donné sur une période donnée, il n'y a aucune raison a priori de pouvoir l'appliquer à un autre peuplement, ni même au même peuplement à un autre moment de sa vie : les probabilités de transition obtenues dépendent bien évidemment de nombreux facteurs parmi lesquels on peut citer la fertilité (essence et station), la densité du peuplement, sa structure et donc le temps si sa structure n'est pas à peu près stationnaire, ..etc. La question se pose donc de savoir comment concevoir et construire véritablement un modèle de transition utilisable.

633 Modèles de transition dans les peuplements hétérogènes

Nous avons vu qu'à chaque peuplement et à chaque intervalle de temps observé correspond une matrice de transition. Ceci étant, comment construire des modèles de transition utilisables pour des peuplements hétérogènes ?

Une première remarque importante s'impose : une matrice de transition est définie pour un intervalle de temps déterminé ; si donc on veut construire un modèle de transition de période Δt , on ne peut utiliser que des données où l'intervalle séparant les deux mesures successives est Δt ou, à la rigueur, $\Delta t/2$ ou $\Delta t/3$ (1).

Une deuxième remarque est la suivante : si l'on veut avoir une chance d'aboutir, il faut entreprendre la modélisation d'un type de peuplement déjà relativement bien défini (région géographique, composition en essences, structure, règles d'exploitation, ...). On peut d'ailleurs noter que ce sont les peuplements de structure stationnaire ou quasi-stationnaire, c'est-à-dire restant à peu près semblables à eux-mêmes à long terme, qui s'accommodent le mieux de ce type de modélisation : il peut donc convenir plus particulièrement à des peuplements "jardinés" soumis à des coupes périodiques effectuées selon des règles de culture à peu près constantes ; la période du modèle de transition sera alors de préférence l'intervalle séparant deux coupes successives.

(1) on peut alors admettre, sous certaines réserves concernant surtout les interventions humaines, que, si la matrice de transition obtenue sur $\Delta t/2$ ou $\Delta t/3$ est T , celle qui correspond à Δt est T^2 ou T^3 .

Supposons donc qu'on étudie un type de peuplement déterminé : on recueille des données, en faisant deux mesures successives, à l'intervalle Δt , d'un certain nombre de peuplements de ce type. A chacun correspond une matrice de transition et ces matrices sont toutes différentes. Deux voies semblent a priori ouvertes pour exploiter ces données :

- 1 Tenter d'exprimer chacune des probabilités de transition observées en fonction de certaines caractéristiques des peuplements : densité, fertilité, temps (âge),...etc. C'est une voie complexe et périlleuse : nous ne nous y engagerons pas.
- 2 Grouper les données par classes de densité \times fertilité \times (âge) \times ...etc, et déterminer une matrice de transition moyenne pour chaque groupe de données ; pour déterminer la matrice moyenne d'un groupe de données, on peut :
 - soit faire la moyenne des matrices de transition observées, ce qui revient à donner le même poids à tous les peuplements observés,
 - soit construire directement une matrice de transition à partir de la somme des données du groupe, de préférence après les avoir toutes rapportées à l'hectare ; si les densités sont peu différentes, cette méthode donne à peu près le même résultat que la précédente, et est plus rapide.

Si le type de peuplement étudié comporte des essences de tempérament et d'intérêt très différents, il faudra en outre les séparer en plusieurs groupes (le moins possible) et rechercher une matrice de transition pour chaque groupe. On voit que l'étude d'un type de peuplement, ou d'une forêt, peut ainsi conduire à un assez grand nombre de matrices de transition différentes.

634 Inconvénients des modèles de transition

En définitive, les modèles de transition présentent plusieurs inconvénients :

- La représentation des influences dynamiques, comme celle de la densité du peuplement sur l'accroissement courant, est difficile et fastidieuse : elle demande l'utilisation d'une matrice de transition distincte pour chaque niveau de la variable "explicative". C'est envisageable quand seule une variable a une influence significative, mais s'il y en a plusieurs le nombre de matrices nécessaire devient vite prohibitif, puisque c'est le produit des nombres de niveaux des variables explicatives retenues.
- La précision de ces modèles est limitée par la nécessité de définir des catégories de grosseur assez larges : sinon, les hypothèses du paragraphe 632.2 ne sont plus vérifiées, et les matrices deviennent en outre trop grosses.
- Ces modèles sont d'une efficacité médiocre en terme de nombre de paramètres nécessaires à la description de la croissance : une matrice 9×9 (81 paramètres) peut ainsi correspondre à une simple fonction "accroissement" à 3 paramètres.
- Cette inefficacité, associée au défaut de représentation des interactions dynamiques, font douter que les modèles de transition puissent nous faire progresser dans la compréhension du processus de la croissance en forêt.

64 MODELES DE CROISSANCE INDIVIDUELLE D'UN ARBRE SOUMIS A LA CONCURRENCE DU PEUPEMENT

641 Définition

Un modèle de croissance individuelle est un modèle dans lequel chaque arbre du peuplement est décrit individuellement par un ensemble de variables, par exemple l'essence, le diamètre, la hauteur, la situation du houppier, ... On peut distinguer deux sortes de modèles de croissance individuelle, selon la façon dont est représentée la concurrence des autres arbres :

- Les modèles de croissance individuelle d'un arbre soumis à la concurrence individuelle de ses voisins ⁽¹⁾ : dans ces modèles, chaque arbre est en outre décrit par sa position (coordonnées x,y), et la concurrence est le résultat des interactions individuelles entre les arbres.
- Les modèles de croissance individuelle d'un arbre soumis à la concurrence du peuplement ⁽²⁾ : dans ces modèles, la position des arbres est ignorée, et la concurrence est le résultat de la pression globale du peuplement.

Ces derniers sont généralement beaucoup plus économiques du point de vue des calculs, mais ils fournissent une représentation moins réaliste et moins détaillée des processus de compétition entre les arbres. Ce sont les seuls que nous aborderons dans ce manuel, car l'expérience montre que les premiers (les modèles de concurrence individuelle), plus complexes et plus exigeants en moyens informatiques, ne présentent pas d'avantages spécifiques pour la prévision de la production.

642 Variables dynamiques et variables allométriques

Dans ces modèles, chaque arbre est donc représenté par un ensemble de variables. Dans les plus simples, les arbres ne sont décrits que par leur diamètre ; mais dans la plupart des cas, ils le sont aussi par leur essence et éventuellement leur hauteur ou la situation de leur houppier dans le couvert. D'autres variables individuelles, comme le diamètre du houppier, le volume ou la biomasse, peuvent être déduites du diamètre et de la hauteur par des relations allométriques.

Nous appellerons variables dynamiques ⁽³⁾ celles qui sont prédites à un instant donné à partir de l'état de l'arbre à un instant précédent, et variables allométriques ⁽⁴⁾ celles qui sont déduites des premières par des relations allométriques indépendantes du temps entre différentes caractéristiques dendrométriques de l'arbre.

643 Modélisation de la concurrence

Le but d'un modèle de croissance individuelle d'un arbre soumis à la concurrence du peuplement est de prévoir l'évolution de chaque arbre sur une période Δt , en fonction des caractéristiques de l'arbre et du peuplement au début de la période : l'état de chaque variable dynamique de l'arbre à la fin de la

(1) en anglais = distance dependant tree model, ou tree position model

(2) en anglais = distance independant tree model

(3) en anglais = dynamic variables

(4) en anglais = allometric variables

période, par exemple le diamètre, est prédit par une équation où figurent notamment la densité du peuplement et le diamètre de l'arbre au début de la période. La densité du peuplement peut être mesurée de diverses façons :

- mesure absolue, comme le nombre de tiges/ha dépassant une certaine grosseur, ou la surface terrière/ha de l'ensemble des arbres ;
- mesure relative, comme le rapport de la surface terrière du peuplement à la surface terrière maximale possible sur la station ;
- tout autre mesure imaginable, comme la biomasse totale des feuilles par unité de surface, quand elle est mesurable ...

Quelle que soit la mesure choisie, elle sera en quelque sorte l'intégrale de variables individuelles des arbres du peuplement, comme la surface terrière du peuplement est la somme des surfaces terrières de ses arbres.

Bien entendu, la densité varie d'un point à un autre du peuplement. Si le modèle est utilisé pour simuler l'évolution d'une placette de grande dimension, de 1 ou 10 ha par exemple, la densité d'ensemble du peuplement risque de masquer des variations locales importantes comme en crée la présence de vides. On peut tourner cette difficulté en divisant la placette simulée en carrés ⁽¹⁾, de 10 × 10 m par exemple, et en considérant que la densité qui affectera les arbres d'un carré est la densité moyenne de ce carré et des huit carrés voisins. Pour les carrés situés au bord de la placette, on peut aller chercher des voisins fictifs à l'autre bout de la placette pour éviter des effets de bordure.

644 Données nécessaires et analyse des données

Pour déterminer les relations dynamiques et allométriques qui figurent dans un tel modèle, on doit disposer de données provenant de placettes d'échantillonnage, permanentes ou temporaires, et de placettes expérimentales. Toutes les variables figurant dans le modèle doivent avoir été effectivement mesurées sur les arbres-échantillons. Si l'on veut définir la concurrence du peuplement en terme de densités locales observées sur de petits carrés (voir § 643 ci-dessus), les mesures doivent être conduites de façon à pouvoir affecter chaque arbre à son carré pour calculer la densité de celui-ci.

Dans les forêts constituées d'essences variées, il se peut fort bien que des essences rares ne soient représentées que par quelques arbres-échantillons répartis sur un petit nombre de placettes. Il vaut mieux alors les regrouper et rechercher pour elles un jeu unique de relations dynamiques et allométriques.

L'ajustement des diverses équations du modèle aux données recueillies sur le terrain demande de l'habileté et une certaine expérience. Les problèmes que cela pose sont abordés dans d'autres chapitres de ce manuel (voir chapitre 5 et annexe A), mais on peut rappeler quelques points importants :

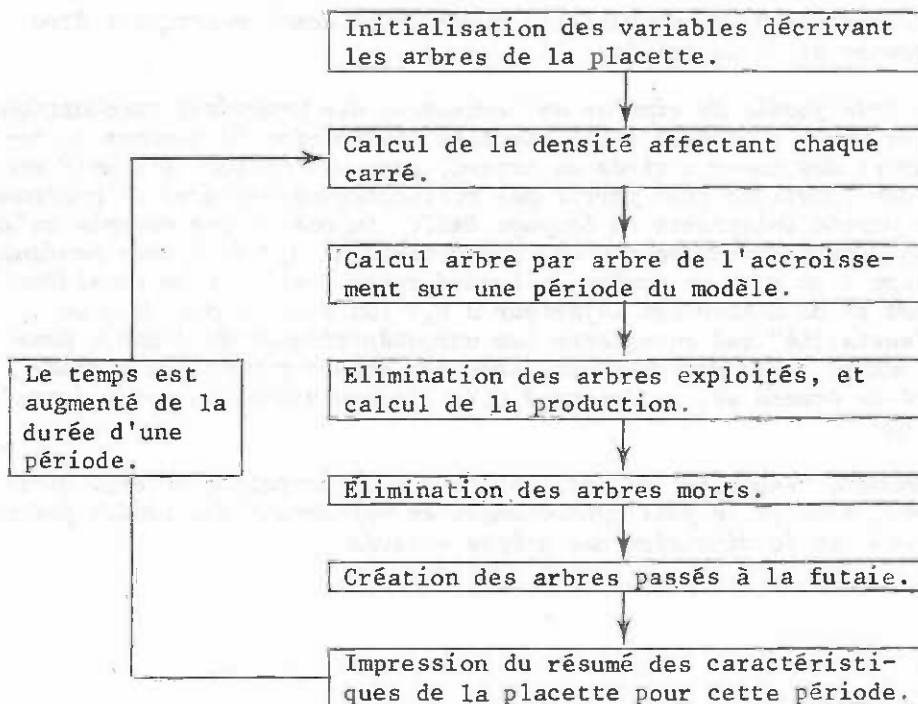
⁽¹⁾ en anglais = quadrats

- Choix d'une fonction qui veuille bien épouser la forme à représenter ;
- Choix d'une technique adéquate d'ajustement, statistique ou graphique, et examen critique des résidus pour dépister les biais ;
- Evaluation de ce qui arrivera si, en utilisant le modèle, on est amené à en extrapoler une relation hors du domaine des observations : le résultat obtenu restera-t-il, même s'il est peu sûr, dans les limites du "raisonnable" ?

645 Structure du modèle

La structure générale du modèle, illustrée par le schéma ci-dessous, est très proche de celle des modèles dynamiques de peuplement présentés au paragraphe 534. Il y a d'abord une phase d'initialisation, dans laquelle on décrit l'état du peuplement (c'est-à-dire de chacun de ses arbres), au début de la période simulée ; cet état peut être fourni par les données d'un inventaire, ou créé de toutes pièces en respectant des distributions fixées a priori.

Le modèle entre ensuite dans une phase cyclique (itérative), chaque itération décrivant une période du modèle, par exemple 1 an ou 5 ans. Chaque itération contient les différentes étapes indiquées sur le schéma, mais leur ordre est sans importance : on peut très bien éliminer les arbres morts ou exploités avant de calculer l'accroissement du peuplement restant.



646 Passage à la futaie, mortalité et exploitation

Du point de vue d'un modèle de croissance individuelle, le passage à la futaie, la mortalité et l'exploitation présentent des points communs :

- ils impliquent la création ou le retrait de certains arbres de la liste des arbres du peuplement,
- du point de vue individuel, ils ont un caractère aléatoire et ne peuvent donc être décrits par de simples relations fonctionnelles,
- on manque souvent de données sérieuses sur ces phénomènes.

Le passage à la futaie et la mortalité sont liés à la densité : les nombres d'arbres passant à la futaie ou mourant pendant une période du modèle pourront s'exprimer en fonction de la densité du peuplement. Une fois ces relations établies, il reste à définir les règles de sélection individuelle permettant d'arriver aux nombres ainsi déterminés.

Pour le passage à la futaie, on peut créer de nouveaux arbres d'essence et de dimension aléatoires, dans le cadre de distributions de probabilité inspirées par des données observées ; la distribution de la dimension peut être à la limite une distribution uniforme : même probabilité pour toute dimension entre la dimension minimum de la futaie et une dimension maximum raisonnable.

Pour la mortalité, la probabilité qu'un arbre donné meurt peut être fonction de l'essence et de la taille.

Il est très facile de simuler en ordinateur des processus stochastiques (aléatoires) de ce genre, à l'aide d'une fonction génératrice de nombres au hasard ⁽¹⁾ qui fournit des nombres tirés au hasard, avec une probabilité uniforme, entre 0 et 1 ; ces fonctions font partie des bibliothèques standard de programme FORTRAN, et sont partie intégrante du langage BASIC. Supposons par exemple qu'un arbre donné de la placette simulée ait une probabilité de 0,7 de mourir pendant la période en cours : on tire un nombre au hasard entre 0 et 1 et on considère l'arbre comme mort si ce nombre est inférieur à 0,7 (on entre alors dans un sous-programme "mortalité" qui enregistre les caractéristiques de l'arbre pour l'impression du résumé et le retire de la liste des arbres vivants du modèle) ; si le nombre tiré au hasard est supérieur à 0,7, l'arbre vivra, au moins jusqu'à la période suivante.

Le phénomène "exploitation" est analogue à la mortalité en ceci qu'il élimine des arbres, mais il le fait généralement en respectant des règles précises concernant l'essence et la dimension des arbres enlevés.

⁽¹⁾ en anglais = random numbers

647 Conclusions sur les modèles de croissance individuelle

Les modèles de croissance individuelle sont extrêmement souples, mais leur construction et leur utilisation exigent une grande compétence en programmation et en analyse des données, et de gros moyens de calcul. Il n'existe pas de méthode stéréotypée de construction de tels modèles, et l'on ne peut que renvoyer le lecteur qui désire approfondir ces questions aux études de référence indiquée dans l'ouvrage bibliographique du C.A.B. (1977) (voir annexe D).

Ces modèles ne constituent pas une panacée. Comme toutes les autres méthodes de prévision, ils supposent l'existence d'une importante base de données constituées de mesures successives sur des placettes permanentes d'échantillonnage et sur des placettes expérimentales soigneusement mesurées et entretenues. D'autre part, comme d'autres méthodes déjà exposées, leur sophistication ne constitue pas une garantie de précision sur l'estimation de la production.



6.1. Conclusions de la partie de travaux effectués

Les travaux effectués ont permis de constater que les modèles de croissance et de production utilisés dans les études de planification sont souvent trop simplifiés et ne tiennent pas compte de la complexité des phénomènes économiques. Il est donc nécessaire de développer des modèles plus réalistes et plus sophistiqués.

Les résultats obtenus montrent que les modèles de croissance et de production peuvent être utilisés pour l'analyse de la dynamique économique et pour l'évaluation des politiques économiques. Cependant, il est important de noter que ces modèles ne peuvent pas être utilisés pour prédire avec précision l'évolution future de l'économie.

CHAPITRE 7

CONTROLE DE LA VALIDITE
DES MODELES
D'ACCROISSEMENT ET DE PRODUCTION

7 CONTROLE DE LA VALIDITE DES MODELES D'ACCROISSEMENT ET DE PRODUCTION

71 NECESSITE DE CE CONTROLE

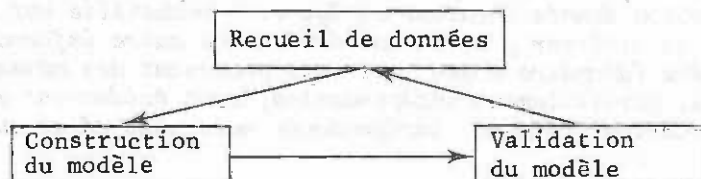
Ce contrôle est nécessaire car c'est lui qui permettra de déterminer dans quelle mesure un modèle représente fidèlement la réalité. Avant d'aller plus loin, notons que :

- Les prévisions faites par un modèle se rapprochent de la réalité, comme une suite d'approximations successives, à mesure qu'augmente l'effort consacré au recueil des données et à la construction du modèle.
- Il arrive un moment où les efforts à consentir pour recueillir de nouvelles données ou construire un modèle statistiquement plus solide deviennent trop coûteux pour l'amélioration marginale que l'on peut en attendre dans la sûreté des prévisions.
- Les modèles forestiers n'offrent pas une précision uniforme sur l'ensemble du domaine qu'ils sont censés couvrir : ils sont plus ou moins précis selon que l'on donne aux variables explicatives une valeur ou une autre.

Le processus de validation d'un modèle peut être aussi long et aussi coûteux que la construction elle-même. On peut être amené à mettre sur pied des expériences dans le but spécifique de le tester, à mettre au point des systèmes spéciaux de traitement des données pour comparer ses prévisions à la réalité d'un grand nombre de placettes, à conduire des analyses statistiques ardues pour étudier les covariances des erreurs résiduelles.

Mais les efforts ainsi consentis ne seront pas perdus. La science diffère de la philosophie et de la religion en ce qu'elle ne peut jamais échapper à la vérification empirique des théories et des modèles qu'elle échafaude : un modèle reste un exercice de spéculation intellectuelle tant que sa validité n'a pas été contrôlée.

En fait, construction et validation d'un modèle sont en interaction perpétuelle dans un processus cyclique que l'on peut schématiser ainsi :



Le contrôle de la validité révèle des faiblesses de comportement du modèle et, ce faisant, permet d'en améliorer la structure ou met en évidence la nécessité de recueillir de nouvelles données sur le terrain. L'expérimentation des modèles, même s'ils ne sont pas valables, peut faire naître l'idée de nouveaux dispositifs expérimentaux ou de nouvelles mesures.

72 LES DONNEES-TESTS

Pour contrôler la validité d'un modèle, on compare ses prévisions à l'observation de situations réelles dont on connaît l'histoire avec précision ; ces données-tests peuvent provenir d'expériences, de placettes d'échantillonnage permanentes ou temporaires, avec l'une ou l'autre des deux éventualités suivantes :

- les données-tests sont également celles qui ont servi à l'ajustement des diverses fonctions du modèle : on parle alors d'auto-validation ;
- les données-tests n'ont servi à l'estimation d'aucun des paramètres du modèle : on parle alors de validation indépendante.

L'auto-validation est quelque chose de très ordinaire : c'est ainsi que l'on juge de la "qualité" d'une régression au moyen de diverses statistiques calculées à partir des écarts entre les valeurs de Y estimées par le modèle d'une part et observées sur les données de l'ajustement d'autre part.

Mais elle peut être dangereusement trompeuse. Prenons toujours l'exemple de la régression ; si les hypothèses du modèle statistique de la régression linéaire (écarts résiduels non corrélés, distribués normalement et de variance uniforme) ne sont pas vérifiées, les statistiques obtenues sont incorrectes (de même, d'ailleurs, que l'estimation des paramètres eux-mêmes). Et, de toutes façons, il est impossible de juger par la seule "qualité" de la régression de la représentativité réelle de l'échantillon sur lequel elle est établie. C'est là la grosse faiblesse de l'auto-validation, qui devient particulièrement dangereuse avec des modèles très complexes ou quand on dispose de peu de données. Avec des modèles simples (disons une ou deux équations) et un grand nombre de données-tests, elle devient plus sûre.

La validation indépendante est un procédé beaucoup plus satisfaisant à tous points de vue car elle permet de juger effectivement de l'efficacité du modèle comme outil de prévision. La meilleure procédure, largement adoptée, est probablement celle qui consiste à séparer les données en deux groupes, par une méthode au hasard, et à utiliser une moitié des données pour la construction et l'autre pour la validation.

La réalité n'est pas toujours aussi tranchée entre auto-validation et validation indépendante : on peut par exemple utiliser pour la construction d'un modèle une information donnée (hauteur et âge ...) recueillie sur un certain nombre de placettes, et utiliser pour sa validation une autre information indépendante de la première (diamètre moyen ...) mais provenant des mêmes placettes. De telles validations, partiellement indépendantes, sont évidemment moins satisfaisantes qu'une validation vraiment indépendante mais plus sûres que l'auto-validation.

73 LES ECARTS RESIDUELS

Le contrôle de la validité des modèles s'appuie généralement sur l'analyse des écarts résiduels (= erreurs résiduelles) , ainsi définis :

écart résiduel = valeur observée - valeur prédite par le modèle

C'est là un concept tout-à-fait analogue à celui de résidu pour la régression et tout ce qui est dit de l'analyse des résidus d'une régression dans l'annexe A26 peut s'appliquer à l'analyse des écarts résiduels d'un modèle.

L'écart résiduel ne donne une information véritable sur les performances d'un modèle que si les valeurs des variables explicatives utilisées par le modèle pour faire sa prédiction sont bien exactement les mêmes que dans l'observation à laquelle on la compare : il faut donc connaître avec précision, sur les données-tests, la valeur des variables explicatives du modèle ; sinon, l'écart résiduel surestime l'erreur du modèle puisqu'il résulte en partie d'une erreur sur les variables explicatives.

Ainsi, par exemple, un modèle de peuplement dynamique assez précis comportera dans ses variables explicatives la densité de plantation, le calendrier et l'intensité des éclaircies : il faut donc disposer, pour contrôler correctement la validité d'un tel modèle, de données-tests comportant ces informations, c'est-à-dire provenant de placettes permanentes suivies pendant longtemps ou d'expériences d'éclaircie.

Un modèle d'accroissement et de production prédit généralement la valeur de plusieurs variables, par exemple diamètre moyen, hauteur dominante, volume total, volume marchand (à une découpe déterminée). Les erreurs relatives résiduelles sur ces variables sont généralement croissantes dans l'ordre : hauteur, diamètre, volume total, volume marchand ; on peut donc souvent se contenter de contrôler la validité du modèle sur le volume total (ou marchand), à moins d'avoir une raison spéciale de connaître la précision des estimations en hauteur ou en diamètre.

On peut résumer de plusieurs façons l'ensemble des écarts résiduels du modèle par rapport aux données-tests :

- En portant sur un graphique les écarts résiduels en fonction des valeurs prédites, ou des valeurs d'une variable explicative (voir alors les commentaires de l'annexe A26).

- En calculant un coefficient de détermination, analogue au R^2 de la régression :

$$1 - \frac{\text{somme des carrés des écarts résiduels}}{\text{somme des carrés des valeurs observées}}$$

- En calculant un écart-type résiduel, analogue à celui de la régression :

$$\sqrt{\frac{\text{somme des carrés des écarts résiduels}}{\text{nombre de données tests} - \text{nombre de variables explicatives} - 1}}$$

- En calculant un écart-type résiduel relatif, analogue à un coefficient de variation :

$$\frac{\text{écart-type résiduel}}{\text{moyenne des valeurs prédites}}$$

S'il est amené à utiliser ou à se référer à ces statistiques, le lecteur devra se souvenir qu'il s'agit là de variables aléatoires dont on ne connaît pas a priori la distribution et que l'on ne peut donc pas relier à des intervalles de confiance ou à des niveaux de significativité spécifiques.

74 COMPARAISONS GRAPHIQUES

L'analyse des écarts résiduels est une technique un peu abstraite. Dans les cas simples, on peut contrôler la validité d'un modèle en portant sur un graphique la variable à laquelle on s'intéresse en fonction d'une variable explicative, d'une part pour le modèle de production et d'autre part pour des peuplements réels. On peut par exemple tracer l'évolution du volume en fonction du temps à la fois pour un modèle et des peuplements réels.

Ce type d'approche permet d'apprécier plus concrètement les forces et les faiblesses d'un modèle que l'analyse des écarts-résiduels, mais il est inutilisable avec des modèles qui mettent en jeu de nombreuses variables explicatives, ou avec de grandes quantités de données-tests.

Nous pensons en fait que les deux types d'approche sont souvent nécessaires : on peut employer l'analyse des résidus avec des masses importantes de données-tests et pour résumer, en quelques statistiques, la précision observée sur l'ensemble du domaine de variation des variables explicatives ; les comparaisons graphiques sont efficaces pour présenter, en publication ou communication, les aspects clés du comportement du modèle.

75 DEFINITION DES LIMITES D'UTILISATION DU MODELE

Les erreurs résiduelles d'un modèle tendent généralement à augmenter au fur et à mesure que l'on va vers des valeurs extrêmes des variables explicatives, surtout si les données-tests couvrent une gamme de stations, d'âges et de conditions de croissance plus vaste que les données utilisées pour construire le modèle.

Il est important, dans le processus de validation, d'examiner en outre comment le modèle se comporte au-delà du domaine couvert par les données-tests en essayant de définir, pour chaque variable explicative, les valeurs au-delà desquelles les prévisions du modèle deviennent franchement fausses. C'est important, car les modèles d'accroissement et de production sont souvent utilisés dans des applications où l'absurdité d'un résultat intermédiaire peut ne pas être immédiatement apparente ; c'est notamment le cas quand ils font partie d'un modèle économique de secteur forestier ou d'un programme de planification des coupes. La validation doit donc définir ces limites d'emploi, à l'intérieur desquelles on peut considérer le modèle comme raisonnablement précis (pour un usage déterminé).

UTILISATION DES MODELES POUR L'AMENAGEMENT

UTILISATION

Les modèles de planification urbaine sont des outils essentiels pour l'aménagement des villes. Ils permettent de visualiser les différents scénarios de développement et d'évaluer leurs impacts sur l'environnement et la société.

Le chapitre 8 est consacré à l'utilisation des modèles pour l'aménagement. Il présente les différents types de modèles et les méthodes de leur utilisation.

CHAPITRE 8

UTILISATION DES MODELES POUR L'AMENAGEMENT

Le chapitre 8 est consacré à l'utilisation des modèles pour l'aménagement. Il présente les différents types de modèles et les méthodes de leur utilisation.

Les modèles de planification urbaine sont des outils essentiels pour l'aménagement des villes. Ils permettent de visualiser les différents scénarios de développement et d'évaluer leurs impacts sur l'environnement et la société.

UTILISATION

Les modèles de planification urbaine sont des outils essentiels pour l'aménagement des villes. Ils permettent de visualiser les différents scénarios de développement et d'évaluer leurs impacts sur l'environnement et la société.

$$M = \frac{1}{1 + \frac{1}{M}}$$

Le chapitre 8 est consacré à l'utilisation des modèles pour l'aménagement. Il présente les différents types de modèles et les méthodes de leur utilisation.

Les modèles de planification urbaine sont des outils essentiels pour l'aménagement des villes. Ils permettent de visualiser les différents scénarios de développement et d'évaluer leurs impacts sur l'environnement et la société.

Le chapitre 8 est consacré à l'utilisation des modèles pour l'aménagement. Il présente les différents types de modèles et les méthodes de leur utilisation.

8 UTILISATION DES MODELES POUR L'AMENAGEMENT

81 INTRODUCTION

Les trois formes principales d'utilisation d'un modèle d'accroissement et de production sont les suivantes :

- (i) table ou jeu de tables, courbe ou jeu de courbes, utilisée telle quelle par les aménagistes ou entrée en ordinateur sous forme tabulée pour la mise à jour de données d'inventaire.
- (ii) programme pour calculateur ou ordinateur, pouvant éditer une table (ou des courbes) de production pour une sylviculture donnée ; cette formule présente un grand intérêt quand le modèle est assez flexible pour qu'il soit impossible de rendre compte dans un seul jeu de tables de toutes les prévisions qu'il permet de faire.
- (iii) sous-programme constituant un sous-modèle dans un programme d'ordinateur pour l'aménagement forestier ; en général celui-ci comprendra aussi une base de données d'inventaire et des contraintes techniques ou économiques sur la sylviculture et l'exploitation.

Les formes d'utilisation (i) et (ii) ont été suffisamment traitées au § 534. Il nous reste à présenter la forme d'utilisation (iii) où le modèle est utilisé conjointement avec des données d'inventaire dans le cadre de l'aménagement forestier ; bien entendu, il faut alors que les variables qui figurent dans le modèle comme variables explicatives figurent aussi dans les données d'inventaire.

82 PEUPLEMENTS EQUIENNES

Lorsque la forêt aménagée est constituée de peuplements équiennes, on peut inclure dans le modèle les trois fonctions suivantes :

$$H_0 = f(E_s, S, A)$$

$$I_g = f(E_s, S, A, G) \quad \text{ou} \quad I_g = f(E_s, H_0, N)$$

$$H_f = f(E_s, H_0, N) \quad \text{ou} \quad H_f = f(E_s, H_0, N, G)$$

dans lesquelles E_s indique l'essence, S la classe ou l'indice de fertilité, H_0 la hauteur dominante, A l'âge, N le nombre de tiges par hectare, G la surface terrière par hectare, I_g l'accroissement en surface terrière par hectare et par an et H_f la "hauteur de forme" définie par $H_f = V/G$ où V est le volume par hectare.

Dans ce cas les variables E_s , A , H_0 , N et G doivent être connues à partir des données d'inventaire, tandis que la classe de fertilité S peut être déterminée au moyen de la première équation.

En général, le modèle est construit à partir d'observations provenant de placettes de production spécialement installées et mesurées pour cela, alors que les données d'inventaire auxquelles on va l'appliquer proviennent des pla-

cettes d'un échantillonnage classique. Si la taille des placettes est la même pour les deux types d'observations, le modèle peut être appliqué sans problème aux données d'inventaire. Si en revanche elle diffère (les placettes d'inventaire sont souvent plus petites que des placettes de production), le modèle peut fournir des résultats biaisés. Il y a à cela deux raisons :

- Si l'on estime la hauteur dominante à partir de la hauteur des n plus gros arbres de la placette de n ares, on risque une légère sous-estimation de H_0 sur les petites placettes (0,01 - 0,02 ha) par rapport aux grandes placettes (0,05 - 0,1 ha), de quelques pour cent au maximum.
- L'effet de la concurrence des arbres extérieurs à la placette augmente quand la taille de la placette diminue (même s'il y a peu d'arbres sur une petite placette, ils peuvent subir une forte concurrence de la part des arbres extérieurs à la placette). Ainsi une fonction où figurent des variables de concurrence ou de densité, comme N et G , ne vaut en principe que pour des placettes de même taille que celles sur lesquelles elle a été établie. Sinon, l'importance du biais éventuel dépend de la différence de taille entre les placettes de production et d'inventaire, du type de fonction d'accroissement incluse dans le modèle et de la variabilité à l'intérieur des peuplements inventoriés : il n'est donc pas possible de donner sur ce point des indications générales.

Lorsque les placettes de production et les placettes d'inventaire ont à peu près la même taille, l'application du modèle aux données d'inventaire doit se faire en prévoyant séparément l'évolution de chaque placette d'inventaire. Quand c'est fait, on peut regrouper les placettes par strates essence \times fertilité \times âge et calculer pour chaque strate les sommes et moyennes des variables faisant l'objet de la prévision.

En revanche, si les placettes d'inventaire sont beaucoup plus petites que les placettes de production, il est souhaitable de commencer par les regrouper par strates essence \times fertilité \times âge et d'appliquer ensuite le modèle à chacune de ces strates plutôt qu'à chaque placette séparément.

83 PEUPELEMENTS HETEROGENES

Quand la forêt aménagée est constituée de peuplements hétérogènes (inéquiennes), on peut inclure dans le modèle de prévision les deux fonctions suivantes :

$$I_v = f(S, P, G_1, G_2, \dots, G_n, N_1, N_2, \dots, N_n, E_1, E_2, \dots, E_n)$$

$$V = f(S, P, G_1, G_2, \dots, G_n, N_1, N_2, \dots, N_n)$$

dans lesquelles :

- I_v est l'accroissement courant en volume marchand (à une découpe déterminée) par hectare et par an ;
- S est la classe de fertilité, définie à partir d'indicateurs du milieu (type de sol, altitude, pluviométrie, ...etc) ;
- P est le type de peuplement ;

- . G_i est la surface terrière/ha du groupe d'essences i , après exploitation ;
- . N_i est le nombre de tiges/ha du groupe d'essences i , après exploitation ;
- . E_i est "l'intensité d'exploitation" pour le groupe d'essences i ;
- . V est le volume marchand/ha, après exploitation.

Bien entendu, là encore, il faut que les variables qui figurent dans le modèle comme variables explicatives soient mesurées dans l'inventaire. Ce que nous avons dit sur la taille des placettes d'inventaire et des placettes de production à propos des peuplements équiennes est aussi valable pour les peuplements hétérogènes.

Malgré tout, il faut bien admettre que l'on en est toujours au stade de la recherche dans le domaine de l'application des modèles d'accroissement aux forêts tropicales hétérogènes, et que l'on dispose de fort peu d'expérience sur le sujet ; nous ne pouvons donc pas donner d'indications précises sur la façon de prévoir l'évolution des forêts tropicales hétérogènes après l'exploitation.

Ce n'est pas le cas pour les forêts hétérogènes de la zone tempérée, où l'on a mis au point des modèles de prévision fiables (voir par exemple MONSERUD, 1980 ⁽¹⁾). Cela vaudrait peut-être la peine d'essayer ces modèles en forêt tropicale hétérogène ; cela n'a jamais été tenté ...

-oO()Oo-

(¹) A. MONSERUD et A. EK, 1980, Comparison of two stand growth models for northern hardwoods - dans H. WRIGHT, 1980 (voir bibliographie en annexe D).



A N N E X E A

.....

TECHNIQUES STATISTIQUES ET MATHÉMATIQUES

UTILISÉES DANS LES ÉTUDES DE PRODUCTION FORESTIÈRE

-o0()0o-

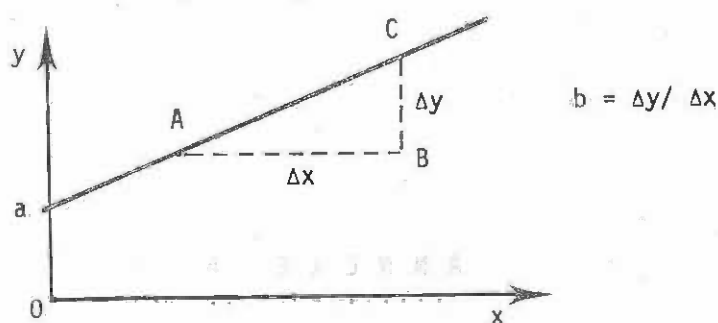


A1 COURBES ET EQUATIONS

Ce chapitre traite des diverses méthodes qui permettent d'exprimer des relations graphiques par des équations.

A11 LA DROITE

La courbe la plus simple est la droite :



On peut exprimer cette relation par l'équation :

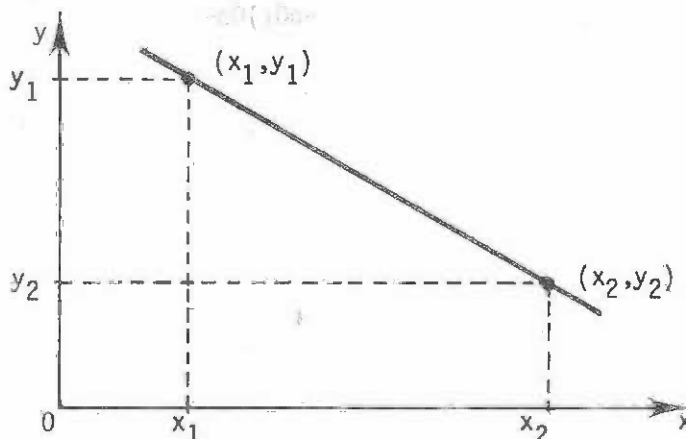
$$y = a + bx$$

a est la valeur de y au point d'intersection de la droite avec l'axe des y (c'est-à-dire pour $x = 0$) ; b est la pente de la droite, qui peut être déterminée comme suit à partir de sa représentation graphique :

- (i) choisir un intervalle Δx approprié (par exemple 10 unités de x) et tracer le segment AB de longueur Δx parallèle à Ox (voir graphique ci-dessus),
- (ii) mesurer la longueur Δy du segment BC parallèle à Oy joignant le point B à la droite ; si C est "au-dessus" de B, Δy est positif ; si C est "au-dessous" de B, Δy est négatif,
- (iii) la valeur de la pente b est donnée par $\Delta y / \Delta x$; remarquons que Δy et Δx doivent être mesurés en unités des axes Oy et Ox , et non en distances réelles sur le papier.

On peut aussi calculer algébriquement les coefficients a et b de l'équation d'une droite :

- (i) choisir sur la droite deux points quelconques de coordonnées (x_1, y_1) et (x_2, y_2) , comme sur l'exemple ci-après :



(ii) la valeur de la pente b est donnée par :

$$b = (y_2 - y_1) / (x_2 - x_1)$$

(iii) la valeur de la constante a est donnée par :

$$a = y_1 - bx_1 \quad (\text{ou } a = y_2 - bx_2)$$

A12 TRANSFORMATION DES VARIABLES POUR L'AJUSTEMENT ⁽¹⁾ D'UNE EQUATION A UNE COURBE

Un certain nombre de relations entre variables concernées dans l'estimation de la production forestière ne prennent pas, en représentation graphique, la forme d'une droite mais d'une courbe quelconque. Si l'on veut obtenir une expression exacte de la courbe, on doit en connaître le modèle, c'est-à-dire l'équation particulière qui la représente. Dans la pratique, on connaît rarement le modèle correct et l'on ne peut en avoir qu'une approximation.

Les équations les plus faciles à ajuster et à manipuler pour procéder à cette approximation sont probablement celles qui ont une forme linéaire, après transformation éventuelle des variables x et y . Les transformations les plus courantes sont :

$\log_{10} x$: logarithme décimal (de base 10)

$\log_e x$ ou $\ln x$: logarithme népérien (de base $e = 2.71828$)

$1/x$: inverse ⁽²⁾

\sqrt{x} : racine carrée

x^2 : carré

On peut faire subir des transformations aux deux variables x et y pour obtenir des courbes de formes diverses.

On peut ainsi essayer, à l'aide de transformations variées, d'obtenir une représentation graphique linéaire de données qui, initialement, ne sont pas liées linéairement. La tâche se trouve simplifiée si l'on a une idée a priori des transformations qui ont une chance de conduire à ce résultat ; nous en avons présenté quelques exemples, pour divers types de données, dans la partie principale de ce manuel.

(1) en anglais = fitting

(2) en anglais = reciprocal.

A13 EXEMPLE D'AJUSTEMENT AVEC TRANSFORMATION DE LA COURBE EN DROITE

Soit une courbe, représentée sur la figure A11 (a), qui traduit une relation entre le volume V d'un arbre et son diamètre D. Nous supposons qu'elle a été tracée à la main et qu'il nous faut maintenant l'exprimer par une équation. On prend des points de la courbe, d'abscisses convenablement espacées, et on dresse un tableau de leurs coordonnées :

Points de la courbe		Transformations		
D	V	D ²	log ₁₀ D	log ₁₀ V
10	0,05	100	1,000	- 1,301
20	0,20	400	1,301	- 0,699
30	0,45	900	1,477	- 0,347
40	0,85	1600	1,602	- 0,071
50	1,50	2500	1,699	0,176
55	1,90	3025	1,740	0,279

On peut voir le résultat d'un premier essai de transformation sur la figure A11 (b), où l'on a représenté V en fonction de D² : les points de la courbe initiale ainsi transformée sont presque alignés, mais il subsiste une légère courbure.

La figure A11 (c) montre le résultat d'une autre transformation : on a pris cette fois le logarithme décimal des deux coordonnées ; la valeur en est donnée, pour les 6 points choisis, dans le tableau ci-dessus. On peut voir que cette transformation permet un ajustement presque parfait par une droite, sauf pour le point de plus petite abscisse qui reste un peu décalé. On peut calculer la pente et la constante de l'équation de cette droite, que nous faisons passer par les deux points (1,301, -0,699) et (1,740, 0,279), signalés par un double cercle sur la figure et correspondant aux valeurs D = 20 et 55 cm de la courbe initiale :

$$b = (0,279 - (-0,699)) / (1,740 - 1,301) = 0,978 / 0,439 = 2,228$$

$$a = -0,699 - 2,228 \times 1,301 = -3,597$$

et l'équation donnant une approximation de la courbe de la figure A11 (a) est :

$$\log_{10} V = -3,597 + 2,228 \times \log_{10} D$$

A14 APPROXIMATION D'UNE COURBE PAR UN POLYNOME

Les équations n'impliquant que deux coefficients sont commodes, quand on recherche l'approximation d'une fonction, car elles peuvent se représenter sous forme de droites après transformations appropriées des variables. Mais il existe des formes de courbes qui ne se laissent pas approcher aussi simplement ; dans ce cas, on obtient souvent de bons résultats en ajustant à la fonction un polynôme ⁽¹⁾ du 2ème ou du 3ème degré.

(1) en anglais = polynomial

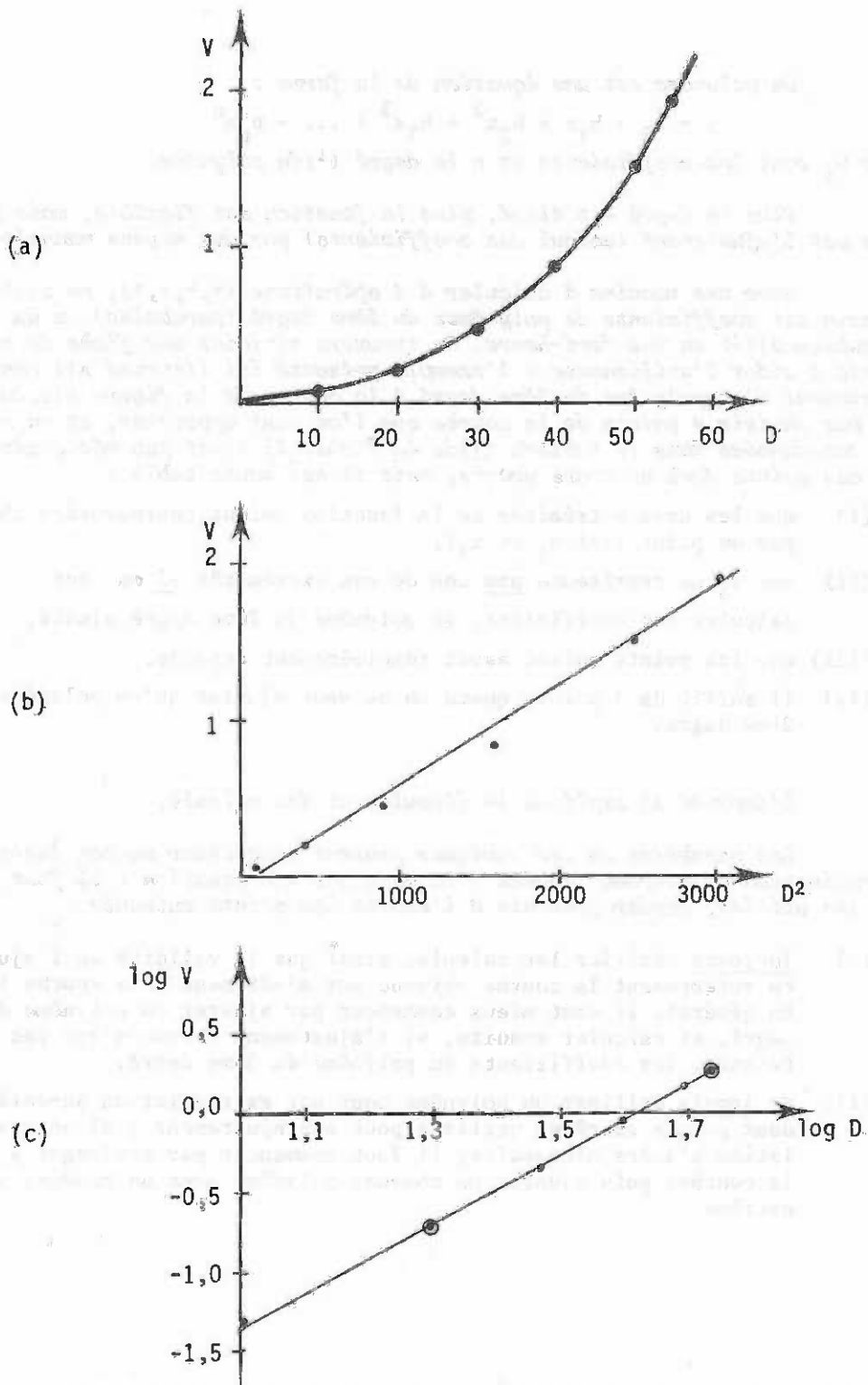


Figure A11 - Transformation en droite, par changements de variables, d'une courbe tracée à la main :

- a) courbe initiale, avec choix des points dont on calculera les coordonnées transformées,
- b) transformation sur x : la nouvelle abscisse est D^2 ,
- c) transformation sur x et y : les nouvelles coordonnées sont $\log_{10} D$ et $\log_{10} V$.

Un polynôme est une équation de la forme :

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n$$

où les b_i sont les coefficients et n , le degré ⁽¹⁾ du polynôme.

Plus le degré est élevé, plus la fonction est flexible, mais plus difficile est l'ajustement (calcul des coefficients) par des moyens manuels.

Avec une machine à calculer à 4 opérations (+, -, ×, ÷), on peut calculer à la main les coefficients de polynômes du 2ème degré (paraboles) ou du 3ème degré (cubiques)⁽²⁾ en une demi-heure. On trouvera ci-joint une fiche de calcul destinée à aider l'utilisateur : l'exemple présenté ici (imprimé A1) montre l'ajustement d'un polynôme du 3ème degré à la courbe de la figure A12. On commence par choisir 4 points de la courbe que l'on veut approcher, et on note leurs coordonnées dans le tableau ① de la fiche. Il n'est pas nécessaire d'inscrire ces points dans un ordre précis, mais il est souhaitable :

- (i) que les deux extrémités de la fonction soient représentées chacune par un point (ici x_1 et x_3),
- (ii) que x_4 ne représente pas une de ces extrémités si on veut calculer les coefficients du polynôme du 2ème degré ajusté,
- (iii) que les points soient assez régulièrement espacés,
- (iv) il suffit de 3 points quand on ne veut ajuster qu'un polynôme du 2ème degré.

L'imprimé A1 explique le déroulement des calculs.

Les paraboles et les cubiques peuvent constituer un bon instrument d'approximation de courbes tracées à la main par une équation ; il faut, lorsqu'on les utilise, garder présents à l'esprit les points suivants :

- (i) Toujours vérifier les calculs, ainsi que la validité de l'ajustement en superposant la courbe obtenue par ajustement à la courbe initiale. En général, il vaut mieux commencer par ajuster un polynôme du 2ème degré, et calculer ensuite, si l'ajustement obtenu n'est pas satisfaisant, les coefficients du polynôme du 3ème degré.
- (ii) Ne jamais utiliser un polynôme pour une extrapolation au-delà des deux points extrêmes utilisés pour son ajustement ; si une extrapolation s'avère nécessaire, il faut commencer par prolonger à la main la courbe, puis ajuster un nouveau polynôme avec un nouveau point extrême.

(1) en anglais = order

(2) en anglais = quadratic polynomial or cubic polynomial

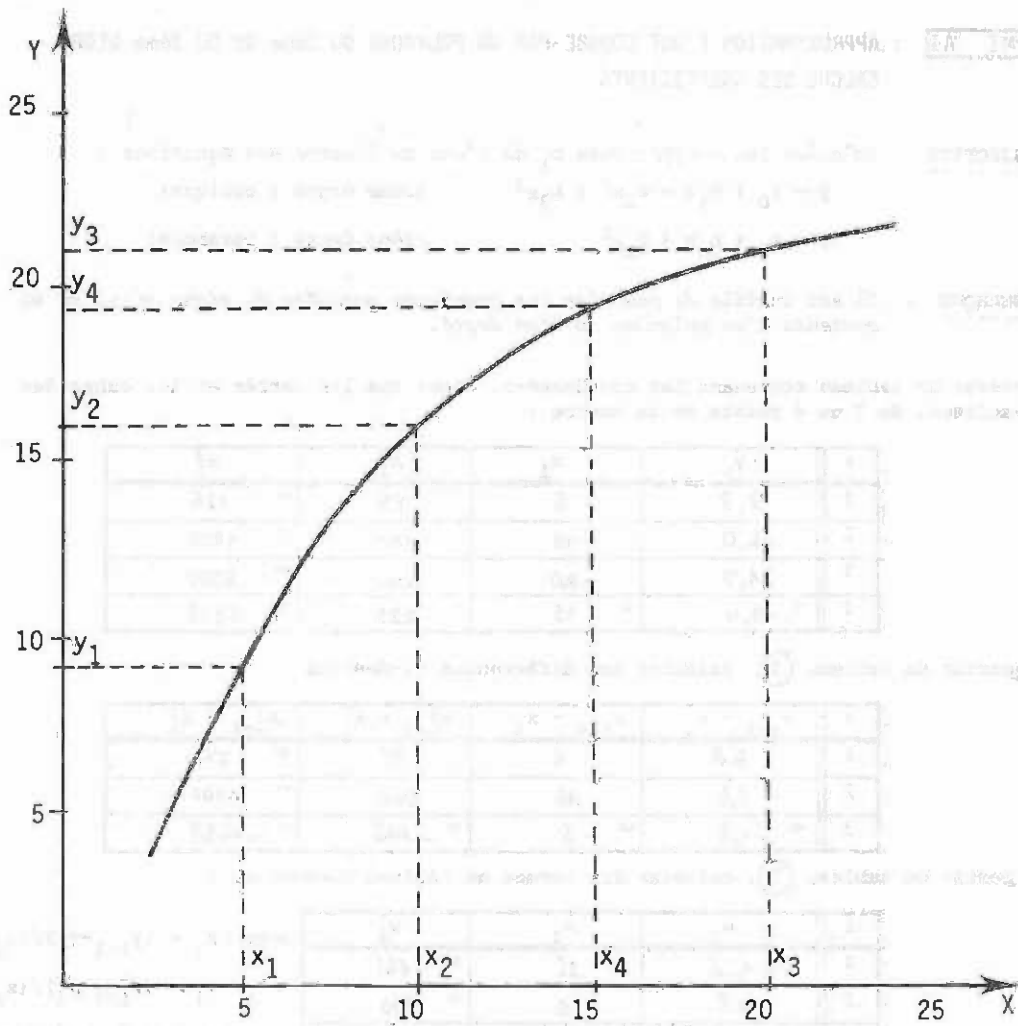


Figure A12 - Courbe tracée à la main et dont on recherche une approximation par une cubique (polynôme du 3ème degré).

(Voir l'imprimé A1 pour les calculs).

IMPRIME A1 : APPROXIMATION D'UNE COURBE PAR UN POLYNOME DU 2ème OU DU 3ème DEGRE
CALCUL DES COEFFICIENTS

OBJECTIFS : Calculer les coefficients b_i de l'une ou l'autre des équations :

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 \quad (3\text{ème degré : cubique})$$

$$y = b_0 + b_1x + b_2x^2 \quad (2\text{ème degré : parabole})$$

REMARQUE : Il est inutile de calculer les grandeurs marquées du signe \diamond si on se contente d'un polynôme du 2ème degré.

- ① Dresser un tableau contenant les coordonnées, ainsi que les carrés et les cubes des abscisses, de 3 ou 4 points de la courbe :

i	y_i	x_i	x_i^2	x_i^3
1	9,2	5	25	\diamond 125
2	16,0	10	100	\diamond 1000
3	24,0	20	400	\diamond 8000
4	\diamond 19,4	\diamond 15	\diamond 225	\diamond 3375

- ② A partir du tableau ①, calculer les différences ci-dessous :

i	$y_{i+1} - y_i$	$x_{i+1} - x_i$	$x_{i+1}^2 - x_i^2$	$x_{i+1}^3 - x_i^3$
1	6,8	5	75	\diamond 875
2	5,0	10	300	\diamond 1000
3	\diamond -1,6	\diamond -5	\diamond -175	\diamond -4625

- ③ A partir du tableau ②, calculer les termes du tableau ci-dessous :

i	u_i	v_i	w_i
1	1,36	15	\diamond 175
2	0,5	30	\diamond 700
3	\diamond 0,32	\diamond 25	\diamond 125

avec : $u_i = (y_{i+1} - y_i) / (x_{i+1} - x_i)$

$v_i = (x_{i+1}^2 - x_i^2) / (x_{i+1} - x_i)$

$w_i = (x_{i+1}^3 - x_i^3) / (x_{i+1} - x_i)$

- ④ A partir du tableau ③, calculer les termes du tableau ci-dessous :

i	$v_{i+1} - v_i$	p_i	q_i
1	15	-0,05433	\diamond 35
2	\diamond 5	\diamond -0,036	\diamond 45

avec : $p_i = (u_{i+1} - u_i) / (v_{i+1} - v_i)$

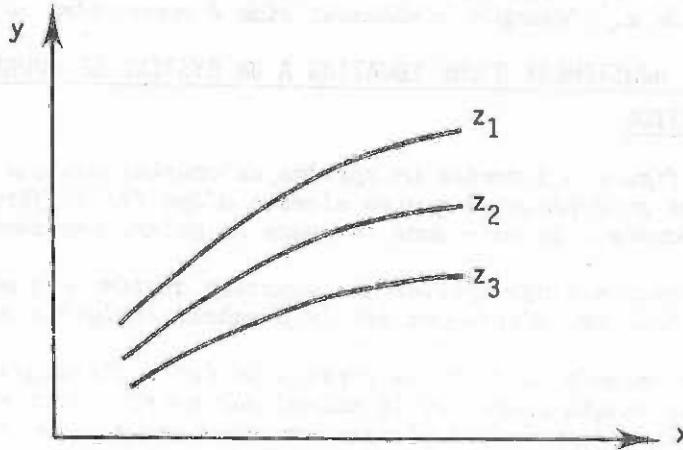
$q_i = (w_{i+1} - w_i) / (v_{i+1} - v_i)$

- ⑤ Calculer les coefficients à l'aide des formules suivantes :

<u>Coefficient</u>	<u>Cubique (3ème degré)</u>	<u>Parabole (2ème degré)</u>
b_3	$(p_2 - p_1) / (q_2 - q_1)$ = 0,002433	0
b_2	$p_1 - b_3q_1$ = -0,1320	p_1 = -0,05433
b_1	$u_1 - b_2v_1 - b_3w_1$ = 2,967	$u_1 - b_2v_1$ = 2,280
b_0	$y_1 - b_1x_1 - b_2x_1^2 - b_3x_1^3$ = -2,599	$y_1 - b_1x_1 - b_2x_1^2$ = -0,466

A15 RELATIONS ENTRE 3 VARIABLES

Jusqu'à maintenant nous avons considéré des fonctions, représentées par des courbes, liant une variable y à une variable x . Mais beaucoup de relations mettent en jeu une troisième variable, que nous appellerons z ; on peut représenter une telle relation de la façon suivante, par exemple :



Ici, on observe pour chaque valeur de z (z_1, z_2, z_3, \dots , etc) une relation distincte entre x et y . Tel est souvent le cas avec les courbes de fertilité où la hauteur (y) varie différemment avec l'âge (x) selon le niveau de fertilité (z), ou encore avec les courbes donnant le volume ou la surface terrière en fonction de la hauteur (x) et de la densité du peuplement (z).

On peut avoir besoin d'exprimer sous forme d'équations un tel système de courbes tracé à la main dans un nuage de points représentant les données, que ce soit pour poursuivre des calculs, pour l'intégrer dans un programme d'ordinateur ou tout simplement pour la commodité de l'analyse et de la communication.

Il y a alors deux façons d'envisager l'expression de ce système :

- (i) soit par une série d'équations distinctes, chacune liant y à x pour une valeur déterminée (un niveau) de z ,
- (ii) soit par une seule équation reliant les 3 variables x, y, z .

La série d'équations distinctes présente deux inconvénients :

- (i) il n'y a pas de moyen de déterminer les valeurs de y pour une valeur de z intermédiaire entre les niveaux retenus.
- (ii) il faut un grand nombre de coefficients pour décrire le système de courbes.

On doit donc donner la préférence à la seconde méthode proposée, c'est-à-dire l'équation unique : elle est plus concise et permet de calculer y pour n'importe quelle valeur de x et de z (dans le domaine ⁽¹⁾ couvert par

(¹) en anglais = range

le système de courbes).

On peut d'ailleurs passer de la série d'équations distinctes, mais de même forme, à l'équation unique en utilisant le procédé d'harmonisation ⁽¹⁾. Le principe en est de représenter graphiquement les valeurs de chacun des coefficients en fonction de z et d'en déduire, pour chaque coefficient b_i , une relation continue $b_i = f_i(z)$ donnant la valeur du coefficient pour n'importe quelle valeur de z. L'exemple ci-dessous aide à comprendre ce procédé.

A16 EXEMPLE : AJUSTEMENT D'UNE EQUATION A UN SYSTEME DE COURBES PAR HARMONISATION

La figure A13 montre un système de courbes hauteur (H)/diamètre (D) pour des arbres appartenant à quatre classes d'âge (A) différentes ; ces courbes ont été tracées à la main dans le nuage de points des données.

On commence par ajuster une équation distincte à chaque courbe ; la forme retenue pour cet ajustement est la parabole (polynôme du 2ème degré).

Les calculs ne sont pas reproduits ici : ils comportent le choix de 3 points sur chaque courbe et le calcul des coefficients suivant la méthode indiquée sur l'imprimé A1. Le tableau ci-dessous donne les valeurs des coefficients obtenues pour chaque classe d'âge :

Classe d'âge	Coefficients		
	b_0	b_1	b_2
5	2,2	1,0625	-0,0219
10	6,6	1,0500	-0,0250
15	9,8	1,1375	-0,0281
20	12,5	1,2750	-0,0313

Pour réduire cette série de quatre équations à une seule, on doit exprimer chaque coefficient b_i en fonction de l'âge. La meilleure façon de se rendre compte de la forme de leur relation à l'âge est de les représenter graphiquement en fonction de l'âge (voir figure A14).

Cette figure montre que cette relation se traduit pour b_0 par une courbe légèrement convexe, pour b_1 par une courbe asymétrique de courbure plus importante, et pour b_2 presque exactement par une droite. Les équations correspondant à ces trois courbes ont été ajustées, en utilisant à nouveau l'imprimé A1 :

$$b_0 = -0,0187 + 1,153 \times A - 3,100 \times A^2 \quad (1)$$

$$b_1 = 1,225 - 0,0508 \times A + 0,00400 \times A^2 - 0,0000667 \times A^3 \quad (2)$$

$$b_2 = -0,0188 - 0,000627 \times A \quad (3)$$

⁽¹⁾ en anglais = harmonization

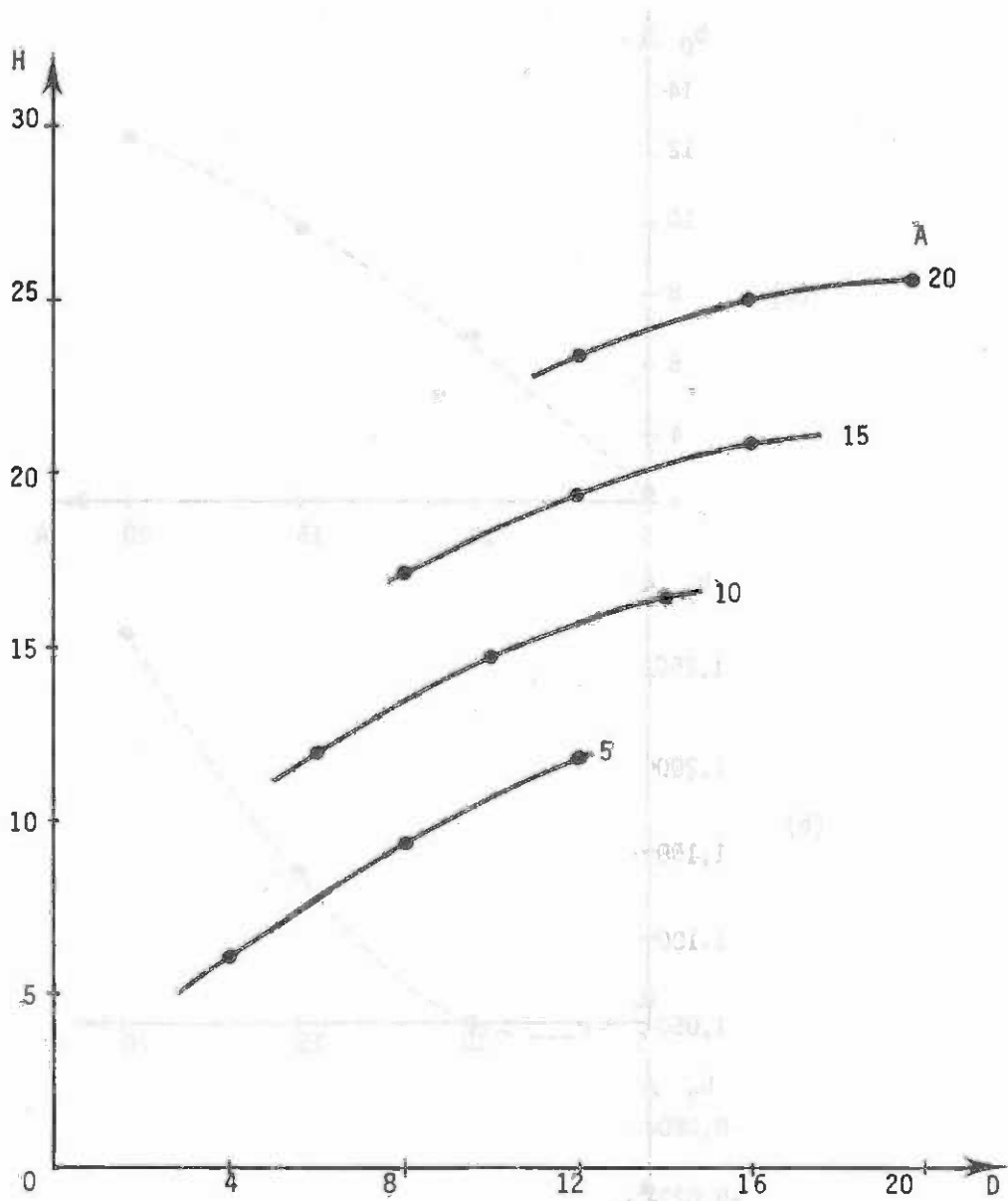


Figure A13 - Courbes hauteur/diamètre tracées à la main pour différentes classes d'âge. On cherche une expression approchée du système par harmonisation de polynômes du 2ème degré.

(Voir dans le texte le traitement de cet exemple).

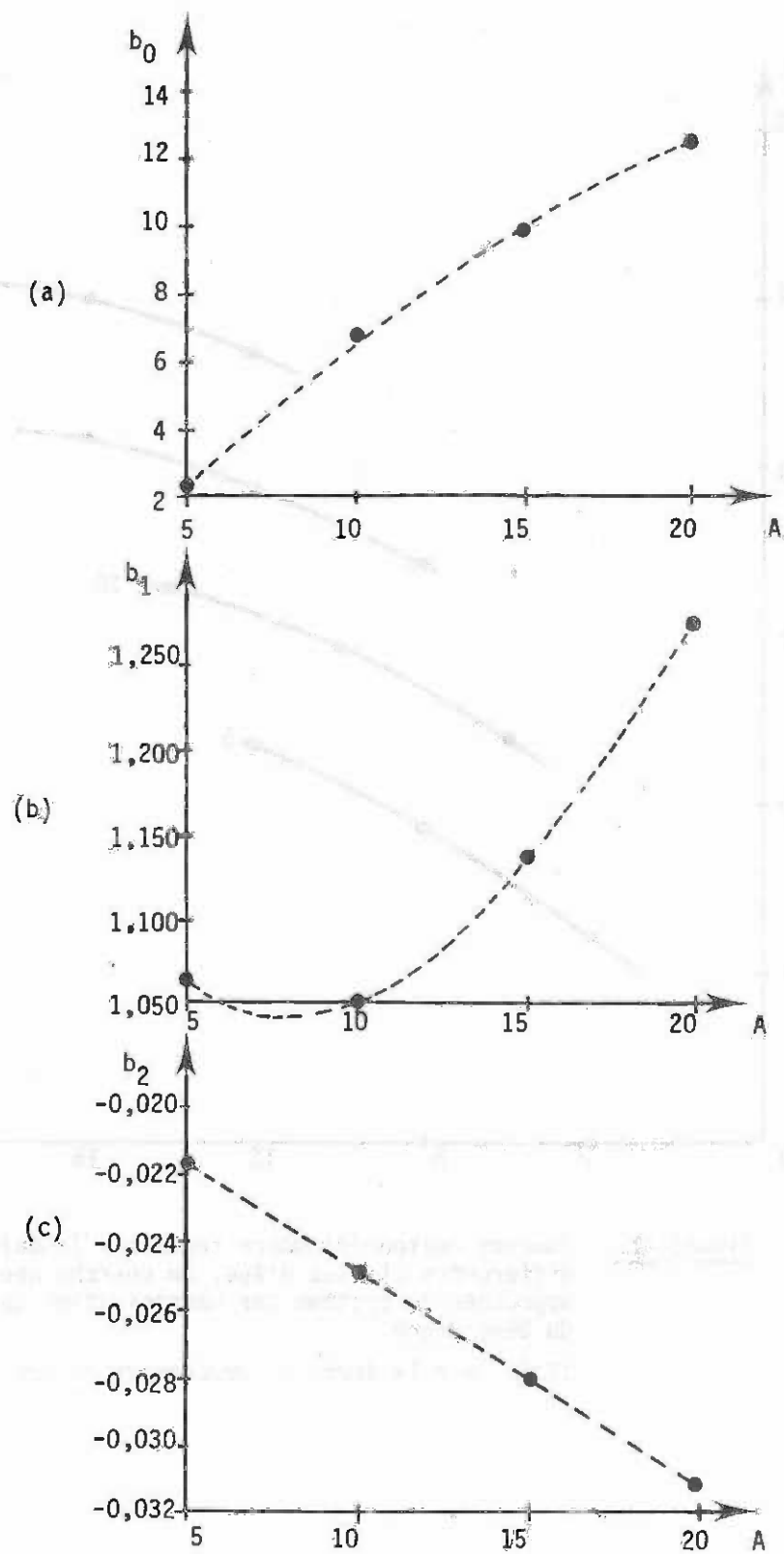


Figure A14 - Valeurs, en fonction de l'âge A, des coefficients des quatre polynômes du second degré ajustés aux quatre courbes de la figure A13.

L'équation générale des courbes de la figure A13 est de la forme préalablement retenue (parabole) :

$$H = b_0 + b_1 D + b_2 D^2 \quad (4)$$

où les valeurs des coefficients b_0, b_1, b_2 peuvent être calculées pour n'importe quelle valeur de l'âge entre 5 et 20 ans, à l'aide des équations (1) à (3). On peut d'ailleurs écrire directement la fonction harmonisée :

$$\begin{aligned} H(D, A) = & - 0,0187 + 1,153 \times A - 3,100 \times A^2 \\ & + (1,225 - 0,0508 \times A + 0,00400 \times A^2 - 0,0000667 \times A^3) \times D \\ & - (0,0188 + 0,000627 \times A) \times D^2 \end{aligned}$$

On a choisi là un exemple d'harmonisation assez complexe pour illustrer toute la portée des principes mis en jeu ; dans la pratique, on peut souvent se permettre certaines simplifications. Ainsi, dans l'exemple ci-dessus, on obtient une assez bonne approximation du faisceau de courbes tracé à la main en donnant à b_1 et b_2 leurs valeurs moyennes, b_0 seul variant linéairement avec l'âge selon une équation :

$$b_0 = a_0 + a_1 A$$

L'équation résultante serait alors :

$$H(A, D) = a_0 + a_1 A + b_1 D + b_2 D^2$$

où a_0, a_1, b_1, b_2 sont des constantes.

Ce modèle ne donne pas un ajustement aussi bon que celui que nous avons développé plus haut, mais il permet de prédire la hauteur avec une précision de $\pm 0,50$ m dans le domaine couvert par les courbes tracées à la main ; ceci reste très acceptable si les données initiales étaient assez dispersées ⁽¹⁾ et donc les courbes tracées à la main elles-mêmes plus ou moins précises.

Comme pour les polynômes, on ne doit jamais faire d'extrapolation à partir de fonctions harmonisées. Si une extrapolation s'avère nécessaire, il faut alors prolonger les courbes à la main sur le domaine où l'on en a besoin et calculer à nouveau tous les coefficients de la fonction d'approximation.

⁽¹⁾ en anglais = scattered

A2 AJUSTEMENT D'UNE EQUATION PAR LA METHODE DES MOINDRES CARRÉS

A21 REGRESSION LINEAIRE SIMPLE

Nombre d'ouvrages exposent en détail, avec de nombreux exemples à l'appui, le principe de la régression linéaire ⁽¹⁾ (voir SNEDECOR, DRAPER et SMITH, SEBER dans la bibliographie). Nous nous contenterons donc ici de donner les très grandes lignes des principes fondamentaux, des calculs de base et des aspects statistiques les plus importants.

La régression linéaire est le plus souvent obtenue par la méthode des moindres carrés ⁽²⁾ : cette technique d'ajustement d'une équation explicative à des données brutes (c'est-à-dire à des observations non lissées) est basée sur la minimisation de la somme des carrés des écarts ⁽³⁾ par rapport à une droite tracée dans le nuage de points.

Lorsque certaines hypothèses, énumérées ci-dessous, sont vérifiées, les paramètres (coefficients) ainsi obtenus pour l'équation ajustée sont ceux qui ont la plus grande probabilité d'être corrects d'un point de vue statistique. Lorsque ces hypothèses ne sont pas toutes vérifiées, la méthode fournit encore une estimation des paramètres, mais ce n'est plus la meilleure estimation possible : on peut en obtenir une meilleure en appliquant le principe du "maximum de vraisemblance" ⁽⁴⁾ au problème posé, mais ceci sort du cadre de la présente annexe.

La figure A21 va nous permettre de présenter le type le plus simple de problème de régression, impliquant une relation entre deux variables x et y . On suppose que la variable x (variable explicative) ⁽⁵⁾ doit permettre de prédire la valeur de y (variable expliquée) ⁽⁶⁾ ; le modèle statistique est :

$$y_i = a + bx_i + e_i \quad (1)$$

où les e_i sont les écarts aléatoires des points i à la droite de régression. Sur la figure A21, la ligne en trait continu représente :

$$y = a + bx$$

que l'on appelle généralement l'équation de régression ⁽⁷⁾ ; a et b en sont les coefficients, dont l'estimation à partir des données est le but essentiel des calculs. Chaque "résidu" ⁽⁸⁾ e_i représente la distance, mesurée verticalement, du point i à la droite de régression.

⁽¹⁾ en anglais = linear regression

⁽²⁾ en anglais = least squares method

⁽³⁾ en anglais = deviations

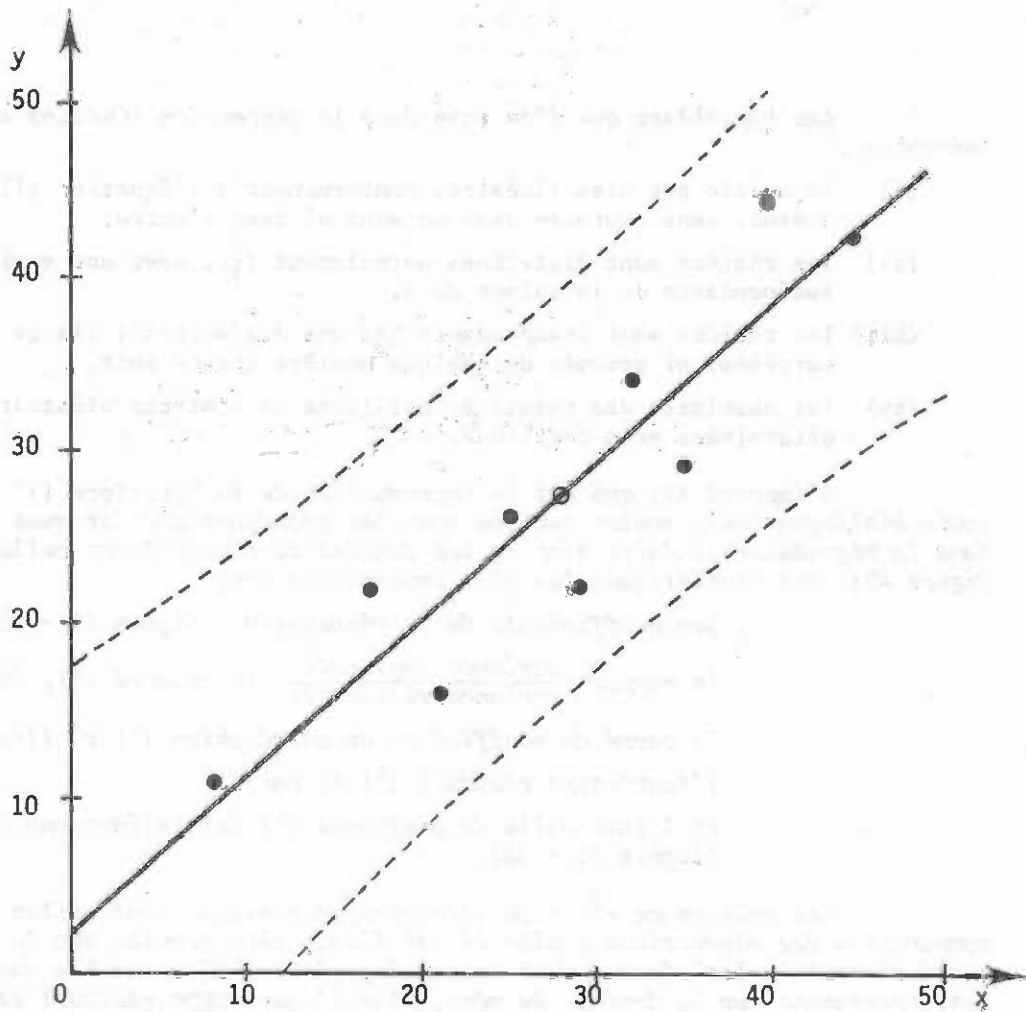
⁽⁴⁾ en anglais = maximum likelihood

⁽⁵⁾ en anglais = predictor variable, ou independant variable

⁽⁶⁾ en anglais = dependant variable

⁽⁷⁾ en anglais = regression equation

⁽⁸⁾ en anglais = residual



- Droite de régression $y = a + bx$
- - - Limite de l'intervalle de confiance, au niveau 95%, de l'estimation par le modèle de la valeur de y correspondant à une valeur de x .
- ⊙ Centre de gravité du nuage, ayant pour coordonnées les valeurs moyennes de x et y .

Figure A21 - Ajustement à un nuage de données du modèle de la régression linéaire simple.

Les hypothèses que l'on fait dans la régression linéaire sont les suivantes :

- (i) le modèle est bien linéaire, conformément à l'équation (1) ci-dessus, sans courbure dans un sens ni dans l'autre,
- (ii) les résidus sont distribués normalement ⁽¹⁾, avec une variance ⁽²⁾ indépendante de la valeur de x ,
- (iii) les résidus sont indépendants les uns des autres : ils ne sont ni corrélés, ni groupés de quelque manière que ce soit,
- (iv) les abscisses des points échantillons ne sont pas aléatoires, mais déterminées avec certitude.

L'imprimé A2, qui est la reproduction de la "Statform 11" de DAWKINS (voir bibliographie), montre quelles sont les grandeurs statistiques à calculer dans la régression linéaire simple, les données de départ étant celles de la figure A21. Les statistiques les plus importantes sont :

- . les coefficients de la régression (lignes 11 - 12)
- . le rapport $\frac{\text{variance expliquée}}{\text{variance résiduelle}}$ (F observé ⁽³⁾, ligne 18)
- . le carré du coefficient de corrélation ⁽⁴⁾ r^2 (ligne 22)
- . l'écart-type résiduel ⁽⁵⁾ (ligne 25)
- . et l'intervalle de confiance ⁽⁶⁾ des estimations de y (lignes 32 - 35).

Les valeurs de r^2 et de l'écart-type résiduel sont utiles pour la comparaison des régressions : plus r^2 est élevé, plus précise est la relation ; quand r^2 vaut 1, les résidus sont tous nuls, c'est-à-dire que les données tombent exactement sur la droite. De même, plus l'écart-type résiduel est proche de zéro, plus grande est la précision avec laquelle la régression permet d'estimer les valeurs de y . L'intervalle de confiance de la régression donne les limites à l'intérieur desquelles se trouve, avec la probabilité 95 %, la vraie moyenne de m valeurs de y estimées par la régression à partir de m valeurs de x ; on a tracé sur la figure A21 les limites de l'intervalle de confiance pour $m = 1$, c'est-à-dire pour l'estimation par la régression de la valeur de y correspondant à une valeur de x .

Remarquons que la droite de régression passe toujours par le point représentant la moyenne des valeurs observées de x et de y . Les limites de l'intervalle de confiance d'une estimation sont courbes et s'écartent rapidement quand on sort du domaine des données qui ont servi à calculer la régression.

(1) en anglais = normally, [au sens de la "loi normale"]
(2) en anglais = variance
(3) en anglais = variance ratio
(4) en anglais = correlation coefficient
(5) en anglais = residual standard deviation
(6) en anglais = confidence limits

IMPRIME A2 : CALCUL DE DIVERSES STATISTIQUES DANS LA REGRESSION LINEAIRE SIMPLE, EN UTILISANT UN IMPRIME DE DAWKINS (1968) REPRODUIT ICI AVEC LA PERMISSION DE L'AUTEUR. LES DONNEES SONT CELLES DE LA FIGURE A21.

Ligne	REGRESSION LINEAIRE et CORRELATION					
1	Nature des observations : Observations hauteur-diamètre sur des arbres					
2	Variable expliquée y' : Hauteur, en mètres					
3						
4	Variable explicative x : Diamètre, en centimètres					
5						
6	Sommes de carrés (SC) et de produits (SP)		$n = 9$	observations		
7	$(\Sigma y)^2 = 60516$	$\Sigma y \times \Sigma x = 61222$	$(\Sigma x)^2 = 63504$	y	x	
8	$\Sigma y^2 = 7718$	$\Sigma yx = 7839$	$\Sigma x^2 = 8134$	11	8	
9	$(\Sigma y)^2/n = 6724$	$-\Sigma y \times \Sigma x/n = 6888$	$(\Sigma x)^2/n = 7056$	22	14	
10	$SCy = \frac{994}{}$	$SPyx = \frac{951}{}$	$SCx = \frac{1078}{}$	16	21	
11	Régression coefficient: $SPyx/SCx = 0,88219 = b$			22	29	
12	Linéaire constante $\bar{y} - b\bar{x} = 2,63128 = a$			29	35	
13	SC expliquée par la régression = $b \times SPyx = 838,96$			34	32	
14	SC totale = $SCy \cdot \text{nombre total de d.d.l.} = n-1 = 8$			26	25	
15	SC résiduelle et nombre résiduel de d.d.l. = obtenus par différence			42	45	
16	ANALYSE DE VARIANCE			F	F de la	
17	Source de variation	SC	d.d.l.	CM	observé	
18	Régression (expliquée)	838,96	1	838,96	34,9	
19	Résiduelle	155,04	7	22,14	29,25	
20	Totale	994	8		(au seuil $p = 0,1\%$)	
21	Carré du coefficient de corrélation, r^2 ,					
22	SC expliquée par la régression/SC totale = $839 / 994 = 0,844 = r^2$					
23	Coefficient de corrélation linéaire, r,					
24	$r = \sqrt{r^2} = 0,919$ avec le signe de b					
25	Ecart-type résiduel = $\sqrt{CM \text{ résiduel}} = 4,71$			= ET résiduel		
26	Coefficient de variation résiduelle en % = $100 \times ET \text{ résiduel} / \bar{y} = 1,91\% = CV\%$			Σy	Σx	
				246	252	
28	Le t de Student pour n-2 d.d.l. et P=0,05 est : 2,36			\bar{y}	\bar{x}	
				27,3	28	
30	L'écart-type du coefficient b est $\sqrt{CM \text{ résiduel} / SCx} = 0,1433$			= ETb		
31	Son t-observé est b/ETb , ou $\sqrt{F\text{-observé}} = 6,156$			= t-observé		
32	Intervalle de confiance de la moyenne y' de m valeurs de y estimées par la régression à partir de m valeurs de x (de moyenne x') :					
34	$\bar{y}' \pm t \times ET \text{ résiduel} \times \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x' - \bar{x})^2}{SCx}}$					
35						

Dans la régression, on appelle la variable x variable indépendante ou explicative : c'est à partir de sa valeur que les prévisions seront faites. La variable y est appelée variable dépendante ou expliquée : on suppose que sa valeur est contrôlée dans une certaine mesure, par la valeur de la variable explicative.

A22 REGRESSION LINÉAIRE A DEUX VARIABLES EXPLICATIVES

Le modèle de la régression à deux variables explicatives est :

$$Y = b_0 + b_1X + b_2Z$$

où X et Z sont les variables explicatives et Y la variable expliquée. Les hypothèses et les principes généraux sont les mêmes que pour la régression linéaire simple. Bien que plus complexe que dans le cas de la régression simple, le calcul à la main des coefficients et des diverses statistiques de ce modèle est encore possible. C'est un modèle pratique, qui peut s'ajuster à des formes de courbes variées, comme on le verra au §A23.

L'imprimé A3 montre le déroulement des calculs, sur les mêmes données que la figure A21 et l'imprimé A2 (hauteur et diamètre de 9 arbres-échantillons), mais avec l'addition d'une variable supplémentaire, le carré du diamètre ; l'équation à ajuster devient alors :

$$H = b_0 + b_1D + b_2D^2$$

En effectuant tous les calculs de l'imprimé, on obtient la valeur de chaque coefficient :

$$b_0 = 8,31177$$

$$b_1 = 0,371583$$

$$b_2 = 0,00953469$$

Il est important de conduire tous les calculs avec 6 chiffres significatifs au moins, et de ne pas arrondir les faibles nombres (comme g_1 , g_2 , g_3 ici) à quelques décimales seulement : on risque sinon une perte de précision considérable.

Les valeurs obtenues pour F -observé et les t -observés ⁽¹⁾ des coefficients sont plus faibles que pour le modèle de régression simple : c'est que l'addition d'une variable supplémentaire augmente l'incertitude sur l'estimation des paramètres concernant la population d'où est tiré cet échantillon. Ici, aucune des deux valeurs de t obtenues pour les deux coefficients estimés b_1 et b_2 n'est significative ⁽²⁾ : ceci indique que le modèle serait tout aussi efficace pour la prévision si l'on en faisait sortir l'une ou l'autre des deux variables.

(1) en anglais = t -value

(2) en anglais = significant

IMPRIME A3 : CALCUL DES COEFFICIENTS ET DE DIVERSES STATISTIQUES D'UNE REGRESSION LINEAIRE A DEUX VARIABLES EXPLICATIVES

Partie 1

DESCRIPTION DU PROBLEME :

Données Hauteur - Diamètre de l'exemple du paragraphe A21, mais avec $Z = X^2$

(1) DONNEES		
Y	X	Z
11	8	64
22	17	289
16	21	441
22	29	841
29	35	1 225
34	32	1 024
26	25	625
42	45	2 025
44	40	1 600

(4) PRODUITS CENTRES	
$\Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/n =$	994,000
$\Sigma xy = \Sigma XY - \Sigma X \times \Sigma Y/n =$	951,000
$\Sigma zy = \Sigma ZY - \Sigma Z \times \Sigma Y/n =$	52 331,667
$\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/n =$	1 078,000
$\Sigma xz = \Sigma XZ - \Sigma X \times \Sigma Z/n =$	57 716,000
$\Sigma z^2 = \Sigma Z^2 - (\Sigma Z)^2/n =$	3 238 501,556

(5) DETERMINANT ET MULTIPLICATEURS DE GAUSS	
$D = \Sigma x^2 \times \Sigma z^2 - (\Sigma xz)^2 =$	159 968 020,9
$g_1 = \Sigma z^2/D =$	0,020 2447
$g_2 = - \Sigma xz/D =$	-0,000 360 797
$g_3 = \Sigma x^2/D =$	$6,7383 \times 10^{-6}$

(6) COEFFICIENTS DE LA REGRESSION	
$b_1 = g_1 \times \Sigma xy + g_2 \times \Sigma zy =$	0,371583
$b_2 = g_2 \times \Sigma xy + g_3 \times \Sigma zy =$	0,00953469
$b_0 = \bar{Y} - b_1 \times \bar{X} - b_2 \times \bar{Z} =$	8,31177

(7) SOMME DES CARRÉS EXPLIQUÉE PAR LA REGRESSION	
$\Sigma \hat{y}^2 = b_1 \times \Sigma xy + b_2 \times \Sigma zy =$	852,34

(8) SOMME DES CARRÉS RESIDUELLE	
$\Sigma d^2 = \Sigma y^2 - \Sigma \hat{y}^2 =$	141,66

(9) NOMBRE RESIDUEL DE DEGRES DE LIBERTE	
d.d.l. = $n-3 =$	6

(2) SOMMES ET PRODUITS			
ΣY^2 7718	ΣY 246	ΣXY 7839	ΣZY 274 661
	n 9	ΣX 252	ΣZ 8 134
		ΣX^2 8 134	ΣXZ 285 468

(3) MOYENNES	
$\bar{Y} = \Sigma Y/n =$	27,3
$\bar{X} = \Sigma X/n =$	28
$\bar{Z} = \Sigma Z/n =$	903,7

ΣZ^2	10 589 830
--------------	------------

IMPRIME A3 : CALCUL DES COEFFICIENTS ET DE DIVERSES STATISTIQUES D'UNE REGRESSION LINEAIRE A DEUX VARIABLES EXPLICATIVES

Partie 2

(10) CORRELATIONS ENTRE VARIABLES	
$r_{yx} = \Sigma xy / \sqrt{(\Sigma x^2 \times \Sigma y^2)} =$	0,9187
$r_{yz} = \Sigma yz / \sqrt{(\Sigma z^2 \times \Sigma y^2)} =$	0,9224
$r_{xz} = \Sigma xz / \sqrt{(\Sigma x^2 \times \Sigma z^2)} =$	0,9768

(11) COEFFICIENT DE CORRELATION MULTIPLE R_1 & R^2	
$R^2 = \Sigma \hat{y}^2 / \Sigma y^2 =$	0,8575
$R = \sqrt{R^2} =$	0,9260

(12) ANALYSE DE VARIANCE				
	Somme des carrés	d.d.l.	Carré moyen	F-observé
Régression	$\Sigma \hat{y}^2 = 852,34$	2	426,17	18,05
Résiduelle	$\Sigma d^2 = 444,66$	$n-3 = 6$	23,61 = s^2	
Totale	$\Sigma y^2 = 994,00$	$n-1 = 8$		

(13) ECART-TYPE RESIDUEL	
$s = \sqrt{s^2} =$	4,86

(14) ECART-TYPE ET t-OBSERVE DES COEFFICIENTS			
Coefficient	b_1	b_2	
Ecart-type s_b	$s \times \sqrt{g_1} = 0,6915$	$s \times \sqrt{g_3} = 0,01262$	
$t = b/s_b$	0,537	0,756	

(15) INTERVALLES DE CONFIANCE DES ESTIMATIONS	
L'écart-type de la moyenne Y' de m valeurs de Y , estimées par la régression à partir de m valeurs de X et Z (de moyennes X' et Z'), est :	
$s_{Y'} = s \times \sqrt{(1/m + 1/n + g_1 x^2 + 2g_2 xz + g_3 z^2)}$	
avec $x = (X' - \bar{X})$ et $z = (Z' - \bar{Z})$. Pour l'estimation d'un ensemble important d'individus, on peut faire $1/m = 0$; pour l'estimation d'un seul individu, $1/m = 1$.	

A23 TRANSFORMATION DES DONNEES ET AJUSTEMENT D'UNE EQUATION

On peut ajuster certaines fonctions non linéaires avec des techniques de régression linéaire, en faisant subir aux variables expliquée et explicative les transformations appropriées. Quand on envisage de le faire, il est souhaitable de suivre les étapes suivantes :

- (i) Représenter graphiquement les données observées sur papier millimétré simple, et voir si la "ligne moyenne" du nuage présente une courbure.
- (ii) Si oui, choisir une transformation appropriée pour les variables expliquée et/ou explicative et calculer leurs valeurs transformées pour chaque observation.
- (iii) Reporter les données transformées sur papier millimétré simple et examiner le nuage pour voir si sa tendance est maintenant linéaire ; si oui, ajuster une régression linéaire sur ces données transformées.

Pour aider au choix d'une transformation appropriée, la figure A22 présente les formes de quelques fonctions couramment utilisées. Les fonctions (a) à (d) peuvent être ajustées par régression simple. Les fonctions (e) et (f) nécessitent deux variables explicatives : on ne peut donc pas examiner graphiquement les données transformées comme en (iii) ci-dessus.

L'effet de la transformation des données sur les hypothèses de base de la régression linéaire entraîne trois conséquences importantes :

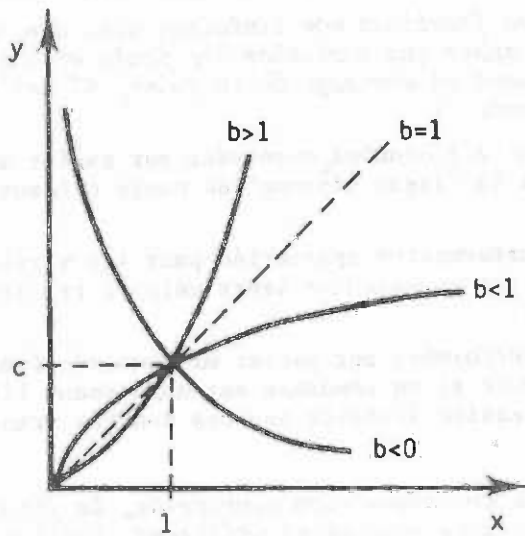
- (i) Des régressions dans lesquelles une même variable expliquée a été soumise à des transformations différentes ne peuvent pas être comparées directement en employant le coefficient de corrélation R (ou R^2) pour juger de la qualité de l'explication.
- (ii) Toute transformation de la variable expliquée modifie la distribution des résidus.
- (iii) Toute transformation de la variable expliquée peut biaiser la régression.

Si l'on veut comparer des régressions du point de vue de la qualité de l'explication quand la variable expliquée a subi des transformations différentes, il faut utiliser l'indice de FURNIVAL ⁽¹⁾ : étant donné un jeu de régressions ayant toutes la même variable expliquée avant transformation, l'équation la mieux ajustée aux données est celle qui a le plus petit indice de FURNIVAL. Cet indice IF se calcule de la façon suivante :

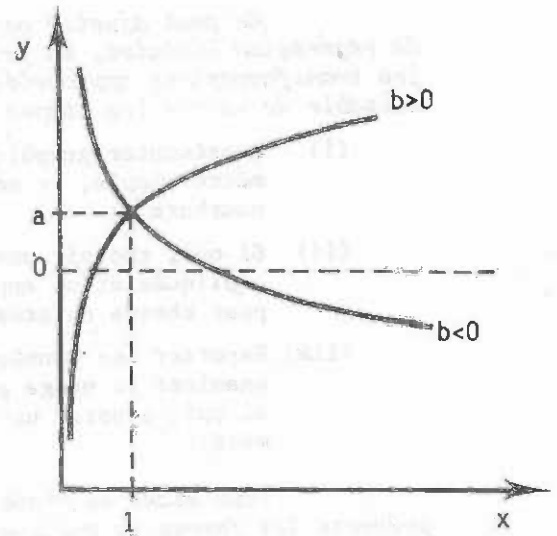
⁽¹⁾ en anglais = FURNIVAL index

[voir G.M. FURNIVAL, 1961 - An index for comparing equations used in constructing volume tables - Forest Science 7(4) 337-341]

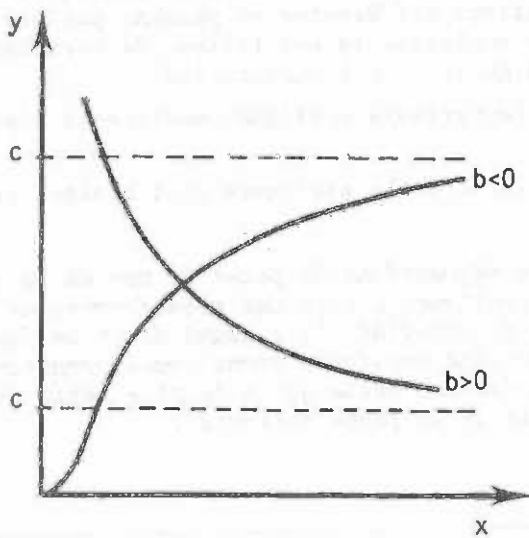
(a) $\log y = a + b \log x$



(b) $y = a + b \log x$



(c) $\log y = a + b/x$



(d) $\log y = a + b \cdot x$

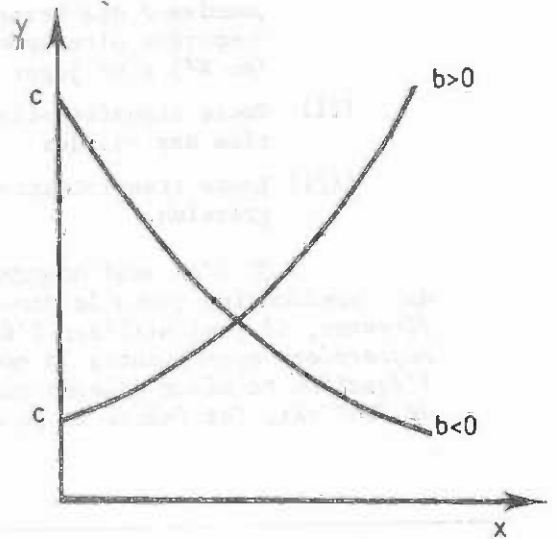
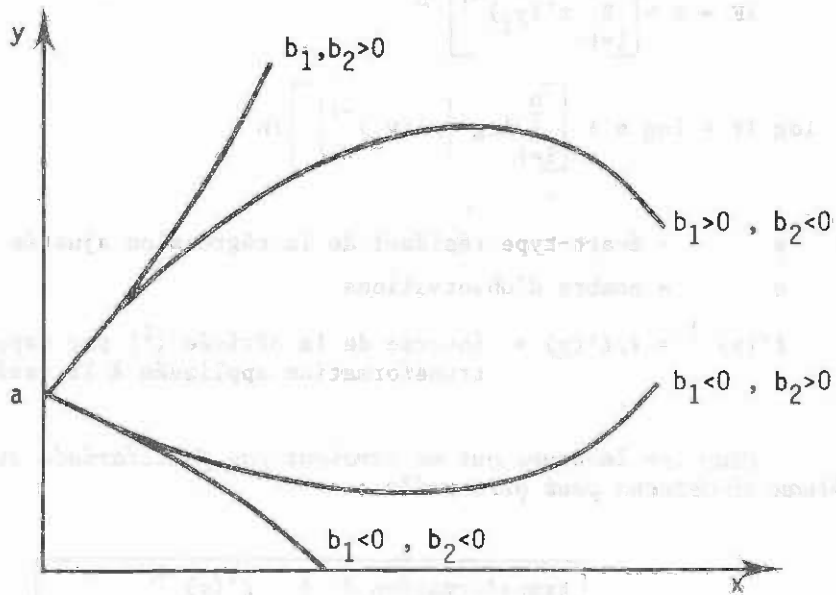


Figure A22 - Formes de courbes engendrées par diverses fonctions ajustables par régression linéaire.

($c = 10^a$ ou e^a suivant que l'on emploie des logarithmes décimaux ou népériens).

(e) $y = a + b_1x + b_2x^2$



(f) $\log y = a + b_1 \log x + b_2x$

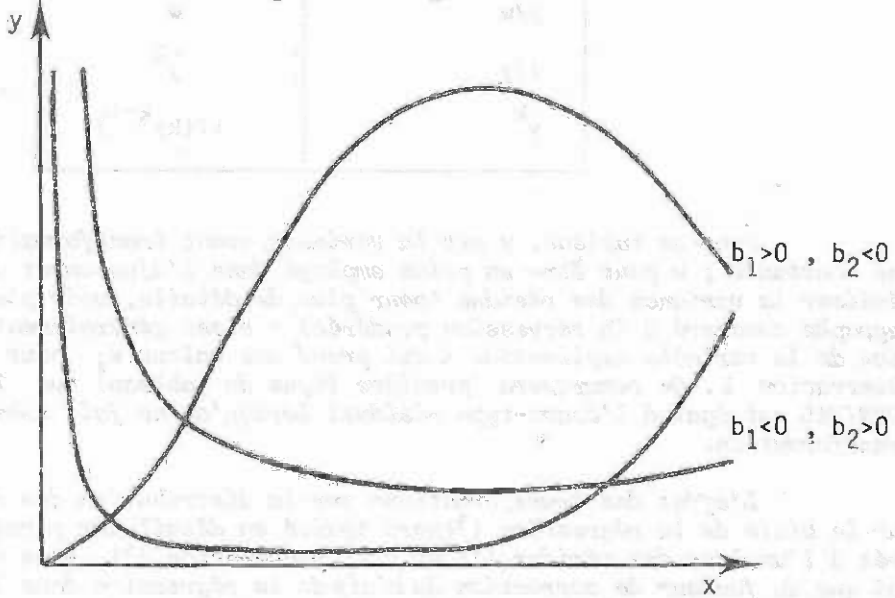


Figure A22 - Formes de courbes engendrées par diverses fonctions ajustables par régression linéaire (Suite)

$$IF = s \times \left[\prod_{i=1}^n f'(y_i)^{-1} \right]^{\frac{1}{n}}$$

ou $\log IF = \log s + \left[\sum_{i=1}^n \log \left(f'(y_i)^{-1} \right) \right] / n$

avec : s = écart-type résiduel de la régression ajustée

n = nombre d'observations

$f'(y)^{-1} = 1/f'(y) =$ inverse de la dérivée ⁽¹⁾ par rapport à y de la transformation appliquée à la variable y .

Pour les lecteurs qui ne seraient pas familiarisés avec les dérivées, le tableau ci-dessous peut être utile :

transformation f	$f'(y)^{-1}$
y	1
$\log_{10} y$	2,3026 y
$\log_e y$	y
y/w	w
$1/y$	$-y^2$
y^k	$1/(ky^{k-1})$

Dans ce tableau, y est la variable avant transformation ; k est une constante ; w peut être un poids employé dans l'ajustement de y pour homogénéiser la variance des résidus (pour plus de détails, voir plus loin le paragraphe consacré à la régression pondérée) : c'est généralement une fonction de la variable explicative x qui prend une valeur w_i pour chaque observation i . On remarquera (première ligne du tableau) que l'indice de FURNIVAL est égal à l'écart-type résiduel lorsqu'on ne fait subir à y aucune transformation.

L'effet des transformations sur la distribution des résidus ⁽²⁾ et sur le biais de la régression ⁽³⁾ sera traité en détail aux paragraphes consacrés à l'analyse des résidus ⁽⁴⁾ et à la pondération ⁽⁵⁾. Nous ne parlerons ici que du facteur de correction du biais de la régression dans le cas d'une transformation logarithmique de la variable expliquée.

(1) en anglais = derivative

(2) en anglais = distribution of residuals

(3) en anglais = bias of the regression

(4) en anglais = residual analysis

(5) en anglais = weighting

Ce facteur de correction du biais, dû à MEYER ⁽¹⁾, suppose que le modèle ajusté est de la forme :

$$\log_a y_i = f(x_i) + e_i$$

où les résidus e_i suivent une même distribution normale d'écart-type s , qui est l'écart-type résiduel calculé dans la régression. Mais le modèle qui sera réellement appliqué est :

$$y = a^{f(x)}$$

où a vaut 10 ou e , suivant que l'on utilise des logarithmes décimaux ou népériens. Si les hypothèses faites sur la distribution des résidus sont exactes, on aura alors une erreur systématique sur y , que l'on peut compenser par l'emploi d'un facteur de correction C :

$$y = C a^{f(x)}$$

Ce facteur de correction est :

(i) $C = 10^{1,1513 s^2}$ pour les logarithmes décimaux (base 10)

(ii) $C = e^{s^2/2}$ pour les logarithmes népériens (base e)

où s est l'écart-type résiduel de la régression de $\log y$.

La validité de ce facteur de correction est liée à la validité des hypothèses faites sur la distribution des résidus. Il est donc impossible d'affirmer a priori que cette correction doit ou ne doit pas être faite dans un cas donné ; il vaut mieux examiner graphiquement la distribution des résidus (voir § A26) avant de prendre une décision.

Dans la pratique, on peut dire que, lorsque le degré d'ajustement obtenu est élevé (soit R^2 supérieur à 0,9), les discussions sur le choix entre les diverses méthodes d'ajustement, la distribution des résidus...etc, sont purement académiques ; la fonction ajustée peut alors être manipulée et transformée comme s'il s'agissait d'une fonction algébrique déterministe.

(1) voir H.A. MEYER (1944) : A correction for systematic error occurring in the application of the logarithmic volume equation - Pennsylvania State University Forest Research paper 7.

A24 LA REGRESSION LINEAIRE MULTIPLE

La régression linéaire simple et la régression linéaire à deux variables explicatives ont respectivement pour modèle :

$$Y = a + bX \quad - (1)$$

et :

$$Y = a + b_1X + b_2Z \quad - (2)$$

où X , Z et Y sont les variables tandis que a, b, b_1, b_2 sont des coefficients à estimer. Ce ne sont en fait que deux cas particuliers du modèle de régression linéaire multiple ⁽¹⁾ :

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_mX_m \quad - (3)$$

ou :

$$Y = b_0 + \sum_{j=1}^m b_jX_j \quad - (4)$$

dans lequel les X_j sont les variables explicatives et les b_j les coefficients à estimer par régression.

Les variables explicatives peuvent être de trois types :

- (i) Chaque variable X_j peut représenter une grandeur physique bien définie ; par exemple, dans un problème donné, X_1 peut être la hauteur et X_2 le diamètre. Ces grandeurs peuvent être plus ou moins corrélées entre elles, mais si cette corrélation en venait à être parfaite ($R = 1$, liaison linéaire entre grandeurs), il faudrait en faire sortir une du modèle.
- (ii) Certaines variables peuvent être des transformations et/ou des combinaisons d'autres variables. Par exemple, on pourrait prendre $X_3 = \log X_2$ et $X_4 = X_1X_2$.

Il faut toutefois, comme nous venons de le dire, s'interdire des combinaisons linéaires du genre $X_3 = X_1 + X_2$ ou $X_1 - X_2$.

- (iii) Certaines, enfin, peuvent être des variables conditionnelles, qui ne peuvent prendre que les valeurs 0 ou 1. Ainsi, dans une régression donnée, X_1 pourrait valoir 0 pour les données observées sur une certaine provenance et 1 pour les données observées sur les autres provenances. Nous en reparlerons dans le paragraphe consacré à la comparaison des régressions et à l'analyse de covariance.

Dans le domaine de la prévision de la production forestière, le nombre de variables explicatives de base est souvent limité à 2 ou 3, avec éventuellement 2 ou 3 autres variables obtenues par transformation ou combinaison des premières : on obtient ainsi des courbes de forme relativement

⁽¹⁾ en anglais = multiple linear regression ou multivariate linear regression

simple. Si on prend comme exemple un modèle de courbes de fertilité :

$$\log H = b_0 + b_1/A + b_2 \times F + b_3 \times F/A$$

on voit qu'il utilise trois variables de base : la hauteur (H) , l'âge (A) et l'indice de fertilité (F) ; ces trois variables sont transformées et combinées pour aboutir au modèle de régression linéaire :

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$$

dans lequel Y est log H , X_1 est $1/A$, X_2 est F et X_3 est F/A.

A25 ESTIMATION DES PARAMETRES DANS L'AJUSTEMENT DU MODELE DE REGRESSION LINEAIRE MULTIPLE

Comme pour la régression linéaire à une ou deux variables explicatives, il faut calculer :

- . les valeurs des coefficients,
- . le coefficient de corrélation multiple R ,
- . l'écart-type résiduel,
- . le F-observé qui est le rapport variance expliquée/variance résiduelle,
- . les écarts-type des coefficients,
- . et les statistiques nécessaires à la construction de l'intervalle de confiance des valeurs de Y estimées par le modèle.

La méthode utilisée est encore celle des moindres carrés.

Ces calculs sont généralement faits par ordinateur, à l'aide d'un programme approprié, car il serait très fastidieux et générateur d'erreurs de les faire à la main. Néanmoins, nous indiquons ici les procédures du calcul manuel à la fois pour ceux qui souhaiteraient les utiliser telles quelles et comme algorithme pouvant servir de base à un programme d'ordinateur, si l'on se trouve dans l'obligation d'en écrire un à partir de zéro.

Suivons donc, étape par étape, le déroulement de ces calculs :

1ère étape : Définition du modèle à ajuster et calcul, s'il y a lieu, de ses variables par transformation ou combinaison.

Le modèle général est :

$$Y_i = b_0 + \sum_{j=1}^m b_j X_{ij} + e_i$$

où Y_i et les X_{ij} représentent une observation i ($i=1$ à n) de la variable expliquée Y et des variables explicatives X_j ($j=1$ à m); les b_j sont les coefficients à estimer et les e_i représentent les résidus (différence entre y_i observé et son estimation par le modèle linéaire).

Pour illustrer ces calculs, nous avons choisi d'ajuster le modèle :

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$$

sur les données :

Y	X ₁	X ₂	X ₃
1	4	9	4
3	3	9	9
7	3	6	6
9	5	8	3
14	4	7	1
18	6	4	5
19	5	3	7
21	7	1	9
24	6	3	4
26	7	2	5

On a donc $n=10$ (nombre d'observation) et $m=3$ (nombre de variables explicatives).

2^{ème} étape : Calcul des sommes et produits croisés, définis comme suit :

(i) Sommes :

$$S_j = \sum_{i=1}^n X_{ij} \quad \text{pour } j = 1, 2, \dots, m$$

$$S_y = \sum_{i=1}^n Y_i$$

(ii) Sommes de produits croisés ⁽¹⁾

$$P X_{jk} = \sum_{i=1}^n X_{ij} \times X_{ik} \quad (\text{comme } P X_{jk} = P X_{kj}, \text{ il suffit de les calculer pour } k \text{ variant de } 1 \text{ à } m, \text{ et } j \text{ variant de } 1 \text{ à } k)$$

$$P Y_j = \sum_{i=1}^n X_{ij} \times Y_i \quad \text{pour } j = 1, 2, \dots, m$$

Avec les données de notre exemple, nous obtenons les valeurs suivantes pour ces grandeurs :

k	y	1	2	3
S _k	142	50	52	53
P Y _k		811	522	752

(1) en anglais = cross-products

$P X_{jk}$:

j \ k	1	2	3
1	270		
2	227	350	
3	267	256	339

Il faut également calculer la somme "totale" des carrés ⁽¹⁾ de Y :

$$\Sigma Y^2 = \sum_{i=1}^n Y_i^2 = 2714$$

3ème étape : Calcul des sommes de produits centrées

A partir des grandeurs calculées à l'étape précédente, on obtient facilement les sommes de produits centrées :

$$Q X_{jk} = P X_{jk} - S_j \times S_k / n$$

$$Q Y_j = P Y_j - S_j \times S_y / n$$

$$\Sigma y^2 = \Sigma Y^2 - S_y^2 / n$$

Avec les données de notre exemple, on obtient :

$Q X_{jk}$:

j \ k	1	2	3	$Q Y_j$
1	20			101
2	- 33	79,6		- 216,4
3	2	- 19,6	58,1	- 0,6

et $\Sigma y^2 = 697,6$

(1) en anglais = total sum of squares

4^{ème} étape : Calcul de la valeur des coefficients et inversion de la matrice des sommes de produits centrées par la méthode d'élimination de GAUSS-JORDAN ⁽¹⁾

Dans un tableau (que nous appellerons C), on dispose les QX et QY de la manière suivante, sans oublier que $Q X_{jk} = Q X_{kj}$:

1	2	3	k	m	1
2							2
⋮							⋮
j				$Q X_{jk}$			$Q Y_j$
⋮							⋮
m							m

Y joue en quelque sorte le rôle d'une $m+1^e$ variable X dans ce tableau de dimensions $(m, m+1)$. Dans notre exemple, le tableau C se présente ainsi :

C_{jk} :

	k	1	2	3	4
j					
1		20	- 33	2	101
2		- 33	79,6	- 19,6	- 216,4
3		2	- 19,6	58,1	- 0,6

Nous pouvons maintenant appliquer à ce tableau C la méthode d'élimination de GAUSS-JORDAN.

METHODE D'ELIMINATION DE GAUSS-JORDAN

p^e colonne

	C_{jk}	C_{jp}	C_{jk}
p^e ligne	C_{pk}	C_{pp}	C_{pk}
	C_{jk}	C_{jp}	C_{jk}

La méthode consiste à pratiquer sur le tableau m "réductions" ⁽²⁾ successives.

⁽¹⁾ en anglais = Jordan elimination process
⁽²⁾ en anglais = reduction

A la p^e réduction, on remplace chaque élément C_{jk} du tableau par une nouvelle valeur C'_{jk} selon l'algorithme suivant :

(i) $C'_{pp} = 1/C_{pp}$ (C_{pp} est le "pivot")

(ii) remplacement de chaque élément C_{jp} de la p^e colonne, sauf le pivot, par C'_{jp} :

$$C'_{jp} = -C_{jp} \times C'_{pp}$$

(iii) remplacement de tous les éléments C_{jk} du tableau, sauf la p^e ligne et la p^e colonne, par C'_{jk} :

$$C'_{jk} = C_{jk} + C'_{jp} \times C_{pk}$$

(iv) remplacement des éléments C_{pk} de la p^e ligne, sauf le pivot, par C'_{pk} :

$$C'_{pk} = C_{pk} \times C'_{pp}$$

Après la dernière (m^e) réduction, les nouveaux éléments C_{11} à C_{mm} constituent la "matrice inverse" de la régression.

Nous verrons que leur connaissance est importante pour le calcul des variances et covariances des coefficients, et donc pour évaluer l'intervalle de confiance d'une estimation par la régression. Les éléments de la $m+1^e$ colonne sont les coefficients de régression b_1 à b_m .

Dans notre exemple, les résultats des trois réductions successives ($m=3$) sont :

REDUCTION 1			
0,05000000	- 1,65000000	0,10000000	5,05000000
1,65000000	25,15000000	- 16,30000000	- 49,75000000
- 0,10000000	- 16,30000000	57,90000000	- 10,70000000
REDUCTION 2			
0,15825049	0,06560636	- 0,96938369	1,78608349
0,06560636	0,03976143	- 0,64811133	- 1,97813121
0,96938369	0,64811133	47,33578528	- 42,94353876
REDUCTION 3			
0,17810238	0,07887895	0,02047887	0,90664807
0,07887895	0,04863523	0,01369178	- 2,56610485
0,02047887	0,01369178	0,02112566	- 0,90721086

Expliquée ainsi, la méthode d'élimination de GAUSS-JORDAN peut paraître complexe ; elle est certainement très fastidieuse si l'on fait les calculs à la main, mais elle est très facilement programmable sur un ordinateur.

On trouvera ci-dessous un extrait de programme en BASIC à partir du tableau C des QX et QY, où les noms des variables sont les mêmes que dans la présentation que nous venons de faire de la méthode :

PROGRAMME EN BASIC (Méthode de GAUSS-JORDAN)

```
100 FOR P=1 TO M
110 LET C(P,P)=1/C(P,P)
120 FOR J=1 TO M
130 IF J=P THEN 190
140 LET C(J,P)=-C(J,P)*C(P,P)
150 FOR K=1 TO M+1
160 IF K=P THEN 180
170 LET C(J,K)=C(J,K)+C(J,P)*C(P,K)
180 NEXT K
190 NEXT J
200 FOR K=1 TO M+1
210 IF K=P THEN 230
220 LET C(P,K)=C(P,K)*C(P,P)
230 NEXT K
240 NEXT P
```

Comme nous l'avons déjà dit, l'exécution de ce programme transforme le tableau C :

- (i) Les éléments C_{11} à C_{mm} constituent maintenant la matrice inverse $Q X^{-1}$ de la matrice des sommes de produits centrées QX.
- (ii) Les éléments $C_{1,m+1}$ à $C_{m,m+1}$ contiennent maintenant les valeurs estimées des coefficients b_1 à b_m .

Dans notre exemple, on obtient :

$$b_1 = 0,9066$$

$$b_2 = -2,566$$

$$b_3 = -0,9072$$

en arrondissant à 4 chiffres significatifs ; on calcule la constante à l'aide de la formule :

$$b_0 = \left(s_y - \sum_{j=1}^m b_j s_j \right) / n$$

soit dans notre exemple :

$$b_0 = \left(142 - (50 \times 0,9066 + 52 \times (-2,566) + 53 \times (-0,9072)) \right) / 10 \\ = 27,82 \text{ (avec 4 chiffres significatifs)}$$

5ème étape : Calcul des trois sommes de carrés = totale (SCT), expliquée par la régression (1) (SCE) et résiduelle (2) (SCR), puis calcul du coefficient de corrélation (R), du F-observé, et de l'écart-type résiduel (s)

Si l'on note :

- Y_i = valeur observée de Y dans l'observation i
- \hat{Y}_i = valeur de Y estimée par la régression à partir des valeurs X_{ij} prises par les variables explicatives dans l'observation i ; le résidu e_i est $Y_i - \hat{Y}_i$
- \bar{Y} = moyenne des Y_i

L'équation ci-dessous montre la définition de chacune de ces sommes de carrés et la relation fondamentale qui les unit :

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i (Y_i - \hat{Y}_i)^2$$

SCT = SCE + SCR

- (i) la somme des carrés totale SCT est $\sum Y^2$, déjà calculée ci-dessus à la 3ème étape
- (ii) la somme des carrés expliquée par la régression SCE est donnée par :

$$SCE = \sum_{j=1}^m b_j \times Q Y_j$$

- (iii) la somme des carrés résiduelle est obtenue en faisant la différence des deux premières :

$$SCR = SCT - SCE$$

Le nombre de degrés de libertés (3) est n-1 pour SCT, m pour SCE et n-m-1 pour SCR.

Le carré moyen (4) expliqué par la régression (CME) est :

$$CME = SCE/m$$

Le carré moyen résiduel (CMR) est :

$$CMR = SCR/(n-m-1)$$

(1) en anglais = regression sum of squares
(2) en anglais = residual sum of squares
(3) en anglais = degrees of freedom
(4) en anglais = mean square.

Le F-observé est :

$F = CME/CMR$ avec m et $n-m-1$ degrés de liberté.

La valeur de F-observé, qui est sous certaines hypothèses une variable de FISHER, sert à tester si la régression est statistiquement significative.

Dans notre exemple, on obtient les valeurs suivantes :

$$SCT = 697,6 \quad (\text{voir } \Sigma y^2 \text{ à la 3ème étape)}$$

$$SCE = 0,9066 \times 101 + (-2,566) \times (-216,4) + (-0,9072) \times (-0,6) \\ = 647,4$$

$$SCR = 697,6 - 647,4 \\ = 50,2$$

$$CME = 647,4/3 \\ = 215,8$$

$$CMR = 50,2/(10-3-1) \\ = 8,367$$

$$F = 215,8/8,367 \\ = 25,79 \quad \text{avec } 3 \text{ et } 6 \text{ degrés de liberté.}$$

Les tables de FISHER indiquent qu'il y a une probabilité $p = 0,1 \%$ que la variable de FISHER à 3 et 6 degrés de liberté dépasse la valeur 23,70 : cela veut dire que l'on aurait moins d'une chance sur mille d'obtenir la valeur de F observée (25,79) s'il n'y avait pas de liaison statistique entre la variable expliquée Y et l'ensemble des variables expliquées X_j ; cette absence de liaison est donc très peu probable.

Le carré du coefficient de corrélation multiple, R^2 , est donné par :

$$R^2 = SCE/SCT$$

ce qui donne, dans notre exemple :

$$R^2 = 647,4/697,6 \\ = 0,9280$$

et : $R = 0,9663$

L'écart-type résiduel s est la racine carrée du carré moyen résiduel :

$$s = \sqrt{CMR}$$

soit dans notre exemple : $s = 2,893$

6ème étape : Calcul des écarts-type et des t-observés des coefficients de la régression.

Les écarts-type des valeurs estimées des coefficients b_1 à b_m sont donnés par la formule :

$$s(b_j) = s \times \sqrt{C_{jj}}$$

et les valeurs de t-observé pour ces mêmes coefficients par :

$$t(b_j) = b_j / s(b_j)$$

On peut tester la significativité de ces valeurs de t-observé qui sont, sous certaines hypothèses, des variables de STUDENT à $n-m-1$ degrés de liberté.

Dans notre exemple, on obtient :

Coefficient	Ecart-type	t-observé	significative
b_1 0,9066	1,221	0,742	-
b_2 -2,566	0,6380	-4,022	**
b_3 -0,9072	0,4205	-2,157	-

Le signe ** indique une significativité au seuil de 1 % (il y a une probabilité $p=1\%$ que la valeur absolue de la variable t de STUDENT à 6 d.d.l. dépasse 3,71 : la probabilité qu'elle atteigne 4,022 est donc encore plus faible). Les coefficients b_1 et b_3 ne sont pas significativement différents de zéro, ce qui ne veut pas dire que l'on puisse les faire sortir tous les deux du modèle de régression. Si l'on éliminait l'un ou l'autre (de préférence le moins significatif, c'est-à-dire b_1), en utilisant la méthode exposée plus loin pour obtenir les nouveaux résultats sans refaire tous les calculs, on verrait que les valeurs de t-observé pour les deux coefficients restants ont augmenté et sont toutes les deux significatives.

7ème étape : Intervalle de confiance d'une estimation par la régression.

L'écart-type d'une estimation de Y faite à partir du modèle pour des valeurs $X_1 \dots X_m$ des m variables explicatives est donné par :

$$s_y = s \times \sqrt{1 + \frac{1}{n} + \sum_{j=1}^m \sum_{k=1}^m C_{jk} x_j x_k}$$

$$x_j = X_j - \bar{X}_j \quad (\bar{X}_j = \text{valeur moyenne de } X_j \text{ sur l'échantillon qui a servi à établir la régression})$$

Supposons, dans notre exemple, que l'on veuille calculer l'écart-type de l'estimation de Y pour $X_1 = 8$, $X_2 = 4$ et $X_3 = 6$.

On commence par calculer les x_j :

$$x_1 = 8 - 5,0 = 3$$

$$x_2 = 4 - 5,2 = -1,2$$

$$x_3 = 6 - 5,3 = 0,7$$

puis la somme $\sum \sum C_{jk} x_j x_k$:

$$\begin{aligned} & 3 \times (3 \times 0,1781 + (-1,2) \times 0,07888 + 0,7 \times 0,02048) \\ & + (-1,2) \times (3 \times 0,07888 + (-1,2) \times 0,04864 + 0,7 \times 0,01369) \\ & + 0,7 \times (3 \times 0,02048 + (-1,2) \times 0,01369 + 0,7 \times 0,02113) \\ & = 1,178 \end{aligned}$$

$$\begin{aligned} \text{enfin } s_y &= 2,893 \times \sqrt{1 + \frac{1}{10} + 1,178} \\ &= 4,366 \end{aligned}$$

L'écart-type d'une estimation de Y à partir de ce modèle pour $X_1 = 4$, $X_2 = 8$ et $X_3 = 6$ est donc 4,366.

Si l'on s'intéresse maintenant à la moyenne de q estimations de Y par le modèle, correspondant à q ensembles de valeurs $X_{\ell 1}, X_{\ell 2}, \dots, X_{\ell m}$ des m variables explicatives ($\ell = 1$ à q), l'écart-type de cette moyenne est donné par une formule à peine plus compliquée :

$$s_{\bar{y}} = s \times \sqrt{\frac{1}{q} + \frac{1}{n} + \sum_{j=1}^m \sum_{k=1}^m C_{jk} x_j x_k}$$

où $x_j = X_j^! - \bar{X}_j$; \bar{X}_j étant toujours la valeur moyenne de X_j sur l'échantillon qui a servi à établir la régression.

$X_j^!$ étant maintenant la valeur moyenne des q valeurs $X_{\ell j}$ de X_j pour lesquelles on estime Y.

Cette formule donne aussi l'écart-type des points de la régression elle-même en faisant $1/q = 0$; en particulier, l'écart-type de la constante b_0 (tous les $X_j^! = 0$) est :

$$s(b_0) = s \sqrt{\frac{1}{n} + \sum_{j=1}^m \sum_{k=1}^m C_{jk} \bar{X}_j \bar{X}_k}$$

Nous avons ainsi exposé les principaux calculs de la régression multiple, mais il nous faut en décrire deux autres, dont on a souvent besoin :

- (i) le calcul de la matrice de corrélation entre variables,
- (ii) le calcul abrégé quand on fait sortir une variable de la régression.

LA MATRICE DE CORRELATION (1)

C'est un tableau qui donne la corrélation entre toutes les variables prises deux à deux. On commence par construire la matrice symétrique V suivante :

	1	2	3	m	m+1
1						
2						
3	Q X					Q Y
⋮						
⋮						
⋮						
m						
m+1	Q Y					Σy^2

On obtient alors très simplement le coefficient de corrélation entre les variables j et k (la variable m+1 est Y) :

$$r_{jk} = v_{jk} / \sqrt{v_{jj} \times v_{kk}} = r_{kj}$$

On peut l'établir dès que l'on a fait la 3ème étape des calculs de la régression.

SORTIE D'UNE VARIABLE DE LA REGRESSION

Quand on a constaté qu'une variable X_r n'était pas significative, on peut souhaiter recalculer les diverses statistiques de la régression amputée de cette variable. Il suffit pour cela de modifier les éléments C_{jk} de la matrice inverse $Q X^{-1}$ et les coefficients b_j de la façon suivante (r est l'indice de la variable indésirable) :

$$b_j' = b_j - C_{jr} \times b_r / C_{rr}$$

$$C_{jk}' = C_{jk} - C_{jr} \times C_{kr} / C_{rr}$$

Dans notre exemple, supposons que l'on veuille faire sortir du modèle la variable X_1 . Avec les 3 variables explicatives, la matrice inverse et les coefficients étaient (avec 6 chiffres significatifs) :

(1) en anglais = correlation matrix

	1	2	3	b
1	0,178102	0,0788790	0,0204789	0,906648
2	0,0788790	0,0486352	0,0136918	-2,56610
3	0,0204789	0,0136918	0,0211257	-0,907211

La sortie de la variable X_1 conduit aux nouvelles valeurs des coefficients :

$$b'_2 = - 2,56610 - 0,0788790 \times 0,906648 / 0,178102$$

$$= - 2,96764$$

$$b'_3 = - 0,907211 - 0,0204789 \times 0,906648 / 0,178102$$

$$= - 1,01146$$

et de la matrice inverse :

	2	3
2	0,0137007	0,00462197
3	0,00462197	0,0187710

On peut ensuite calculer pour la nouvelle régression les diverses autres statistiques (5ème étape et suivantes).

On peut de la même façon éliminer successivement une deuxième variable, puis une troisième, ... etc.

Voici donc présentés brièvement tous les calculs de la régression multiple. On les confie généralement à un ordinateur, mais on peut encore les faire à la main si le problème traité n'est pas trop complexe. Il faut alors veiller, au cours du processus d'inversion de la matrice, à toujours garder un certain nombre de chiffres significatifs (6 au minimum) et donc à ne pas arrondir à un certain nombre de "chiffres après la virgule".

A26 ANALYSE DES RESIDUS

L'analyse des résidus est l'accessoire obligé d'un usage intelligent de la régression multiple. Son objectif est triple :

- (i) vérifier que les résidus de la régression sont conformes aux hypothèses du modèle, c'est-à-dire qu'ils sont distribués normalement ⁽¹⁾, avec une variance homogène, et ne sont pas corrélés entre eux.
- (ii) déceler un éventuel "défaut d'ajustement" ⁽²⁾ du modèle dans l'apparition d'une tendance systématique dans le nuage des résidus.

⁽¹⁾ en anglais = normally distributed (au sens de la loi normale)

⁽²⁾ en anglais = lack of fit

- (iii) étudier visuellement la forme des relations entre les résidus et les variables explicatives que l'on n'a pas encore introduites dans le modèle.

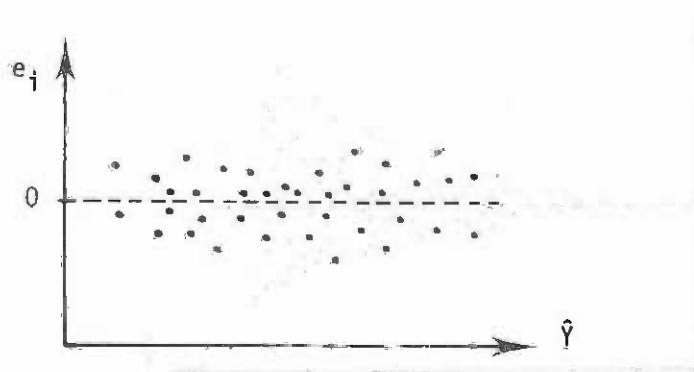
Comme la plupart des calculs de la régression multiple, l'analyse des résidus est très fastidieuse à exécuter à la main, mais elle fait normalement partie, sous une forme ou une autre, de tout bon programme de régression multiple.

Le résidu e d'une observation est la différence entre la vraie valeur Y (valeur observée) de la variable expliquée et la valeur \hat{Y} estimée à partir de la régression :

$$e_i = Y_i - \hat{Y}_i, \text{ pour l'observation } i.$$

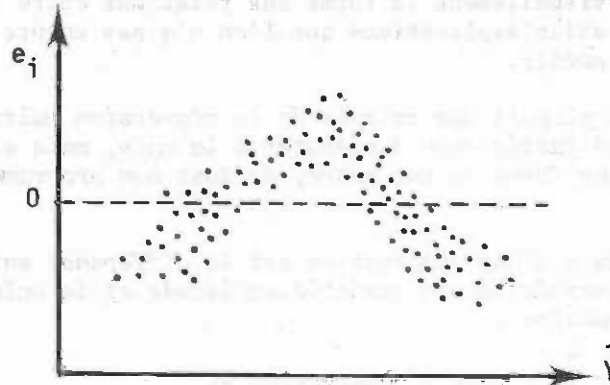
Si on représente graphiquement les e_i en fonction des \hat{Y}_i , on peut obtenir des nuages d'allures variées :

- (i)



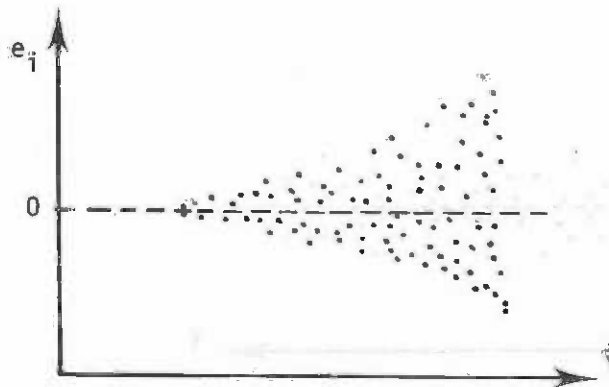
C'est le cas idéal, où les résidus sont uniformément répartis autour de zéro, avec une variance homogène ; le modèle ne présente apparemment pas de caractéristiques indésirables.

(ii)



Cette distribution marque une tendance systématique des e_i en fonction de \hat{Y} , qui révèle un défaut d'ajustement. Autrement dit, on pourrait ajuster aux observations un meilleur modèle en faisant subir aux mêmes variables de base de nouvelles transformations.

(iii)



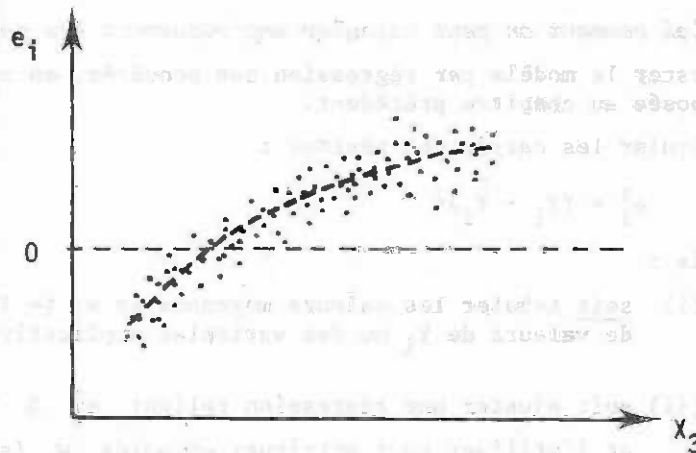
Ici la variance des e_i n'est pas homogène par rapport à \hat{Y} . Dans ce cas, on peut obtenir une meilleure estimation des coefficients en utilisant une transformation de la variable Y ou une régression pondérée (voir § A27).

Si l'on représente graphiquement les e_i en fonction d'une variable X non encore incluse dans le modèle, ils peuvent être distribués au hasard ou présenter une tendance plus ou moins systématique comme en (ii) ci-dessus. Dans ce cas, il serait bon d'inclure la variable X dans le modèle ; si la tendance du nuage des résidus n'est pas linéaire, mais prend l'allure d'une des courbes de la figure A22, on fera subir à X la transformation appropriée avant de l'inclure dans le modèle.

Supposons, par exemple, que l'on ait le modèle :

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

et que le graphique des résidus en fonction d'une troisième variable X_3 présente l'allure suivante :



La tendance moyenne (---) semble être une courbe légèrement convexe, dont on peut probablement rendre compte en ajoutant X_3 et X_3^2 (voir figure A22 (e)) au modèle, qui devient :

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_3^2$$

Après ajustement de ce second modèle, le graphique des résidus en fonction de X_3 aura probablement une allure semblable au premier graphique présenté en (i) ci-dessus.

A27 REGRESSION LINEAIRE PONDEREE

On utilise la régression linéaire pondérée (1) quand les résidus n'ont pas une variance homogène dans le domaine de variation de Y . Ceci se produit fréquemment quand on fait subir à la variable expliquée des transformations complexes ou quand on traite, dans des problèmes d'estimation de la production, des données de volume (où la variance augmente avec le volume).

Dans la régression linéaire pondérée, on associe à chaque observation i un poids w_i judicieusement déterminé, soit empiriquement, soit par des considérations théoriques, et on minimise la somme $\sum_i w_i e_i^2$ (méthode des moindres carrés pondérés (2)). C'est en prenant pour w_i des valeurs proportionnelles à $1/s_i^2$, où s_i est l'écart-type de Y pour la valeur Y_i , qu'on obtient le meilleur ajustement.

(1) en anglais = weighted linear regression

(2) en anglais = weighted least squares

Voici comment on peut calculer empiriquement les poids w_i :

[1] Ajuster le modèle par régression non pondérée, en suivant la méthode exposée au chapitre précédent.

[2] Calculer les carrés des résidus :

$$e_i^2 = (Y_i - \hat{Y}_i)^2$$

[3] Puis :

(i) soit tabuler les valeurs moyennes de w_i ($= 1/e_i^2$) par classe de valeurs de Y_i ou des variables explicatives X_i .

(ii) soit ajuster une régression reliant e_i^2 à Y_i ou aux X_i et l'utiliser pour attribuer un poids w_i ($= 1/e_i^2$) à chaque observation ; une régression simple de e^2 sur une puissance (1,2,3,...) de Y ou de l'une des X peut convenir dans la plupart des cas.

Des considérations théoriques peuvent aussi conduire à une formule donnant s_i et donc w_i en fonction des X_i ou de Y_i .

Pour la régression linéaire pondérée à une seule variable explicative, les calculs à faire sont les suivants :

[1] Calculer les sommes, sommes de carrés, sommes de produits pondérées :

$$A = \sum w_i \quad B = \sum w_i X_i \quad C = \sum w_i X_i^2$$

$$D = \sum w_i Y_i \quad E = \sum w_i Y_i^2 \quad F = \sum w_i X_i Y_i$$

[2] Calculer les moyennes et les sommes de carrés et de produits centrées pondérées :

$$\bar{x} = B/A \quad \bar{y} = D/A$$

$$SC_x = C - B^2/A$$

$$SP_{yx} = F - B \times D/A$$

$$SC_y = E - D^2/A$$

[3] Continuer les calculs à partir de la ligne 11 de l'imprimé A2
(Statform 11 de DAWKINS, 1968).

Pour la régression multiple pondérée, les principes généraux sont les mêmes ; les bons programmes de régression offrent souvent une possibilité de pondération.

Pour justifier l'emploi de la régression pondérée, il faut :

- soit avoir un argument théorique sérieux a priori,
- soit avoir observé une tendance bien nette dans la distribution des carrés des résidus en fonction de \hat{Y} , qui pourra servir de base à une pondération empirique.

Si l'on a obtenu par régression des fonctions bien ajustées, avec un R élevé, il est peu probable qu'une pondération puisse améliorer sensiblement la qualité de l'ajustement. On peut considérer la régression pondérée comme un perfectionnement de la régression, à n'employer qu'après avoir terminé l'analyse classique des données.

A28 COMPARAISON D'EQUATIONS OBTENUES PAR REGRESSION

On a souvent besoin de savoir si plusieurs fonctions obtenues par régression sont suffisamment semblables pour que l'on soit autorisé à les remplacer par une seule. Tel est le cas, par exemple, quand on souhaite regrouper des données d'essences différentes ayant un comportement similaire, ou de forêts, de régions, de types de station ou de provenances différentes.

Les régressions à comparer peuvent être des régressions simples (avec une seule variable explicative) ou multiples. Il existe deux méthodes principales de comparaison :

- (i) L'analyse de covariance : elle convient peut-être mieux au calcul manuel et à la régression simple, bien que rien n'empêche de l'appliquer à la régression multiple.
- (ii) Les tests de significativité sur des variables conditionnelles : cette méthode s'applique très bien à des interactions et à des modèles complexes, mais conduit à une formulation comportant un grand nombre de variables. Elle permet d'adapter des programmes standards de régression multiple à des problèmes de comparaison de régression.

L'analyse de covariance ⁽¹⁾ constitue un vaste sujet. Nous n'en retiendrons ici que les techniques à employer pour comparer les pentes et les constantes de plusieurs régressions simples. Nous avons choisi comme exemple, pour présenter les calculs, les courbes volume-hauteur de deux plantations (I et II) géographiquement distinctes, construites chacune sur l'observation de 10 placettes temporaires ; la figure A23 est la représentation graphique de ces données, présentées dans le tableau A21. Le problème est de décider si l'on peut rassembler les 20 données pour leur ajuster une courbe de régression unique et satisfaisante pour les deux plantations. Le modèle à ajuster est la fonction :

$$\log_{10} V = a + b \times \log_{10} H \quad - (1)$$

(¹) en anglais = covariance analysis

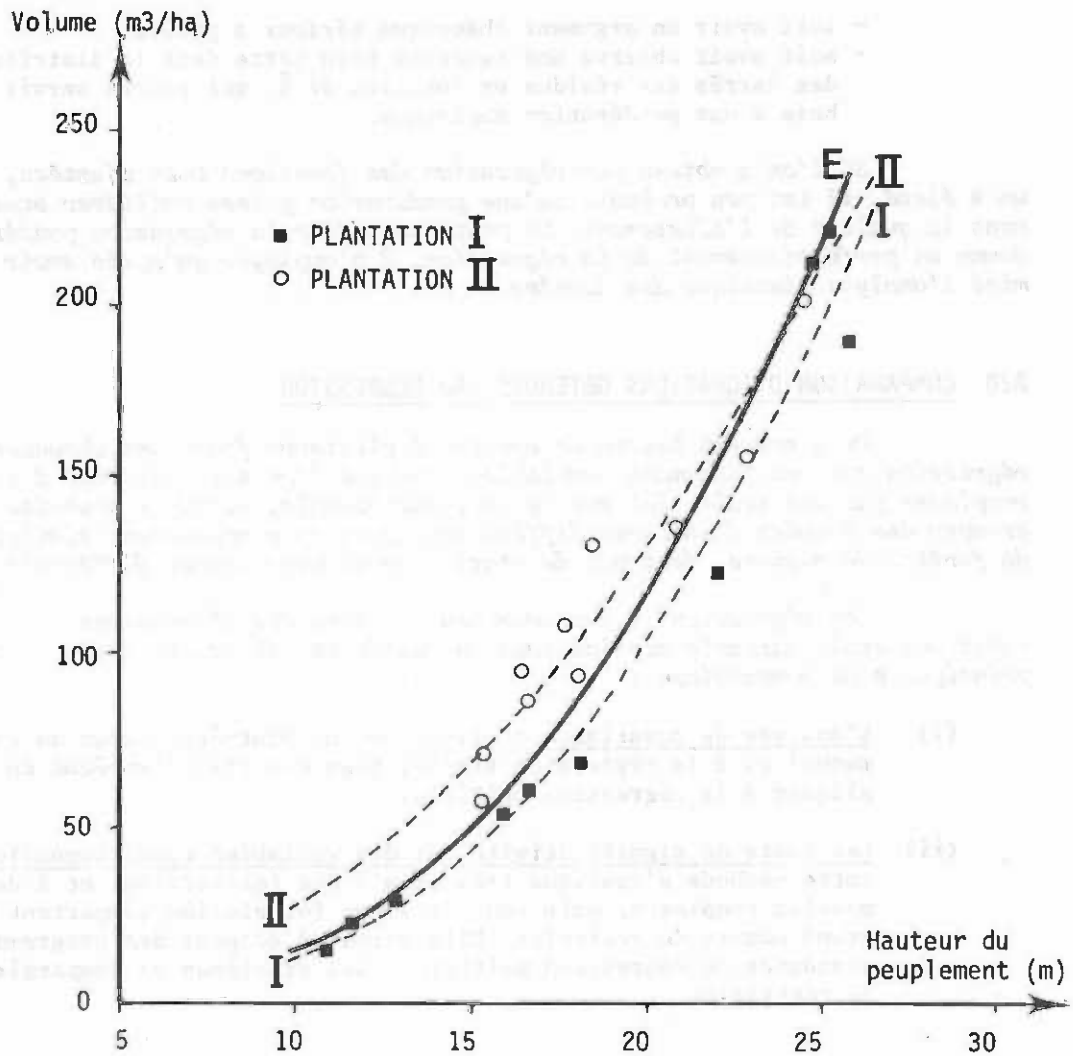


Figure A23 - Exemple de comparaison de courbes de régression volume-hauteur.

(- - - - : courbes de régression de chaque plantation)
(——— : régression d'ensemble)

Tableau A21 : Données, originales et transformées, des deux plantations de l'exemple de comparaison de régressions.

Placette n°	Volume (m3/ha)	Hauteur (m)	Log ₁₀ V y	Log ₁₀ H x
<i>PLANTATION I</i>				
1	30,0	12,9	1,4771	1,1106
2	23,3	11,6	1,3674	1,0645
3	212,6	24,9	2,3276	1,3962
4	15,1	10,8	1,1790	1,0334
5	50,4	15,8	1,7024	1,1987
6	189,4	25,8	2,2774	1,4116
7	68,7	18,2	1,8370	1,2601
8	221,2	25,4	2,3448	1,4048
9	123,9	22,1	2,0931	1,3444
10	60,8	16,7	1,7839	1,2227
<i>PLANTATION II</i>				
11	70,6	15,4	1,8488	1,1875
12	131,3	18,5	2,1183	1,2672
13	201,7	24,6	2,3047	1,3909
14	156,6	22,9	2,1948	1,3598
15	136,2	20,8	2,1342	1,3181
16	87,0	16,7	1,9395	1,2227
17	95,0	16,5	1,9777	1,2175
18	108,4	17,7	2,0350	1,2480
19	93,6	18,0	1,9713	1,2553
20	57,7	15,3	1,7612	1,1847

C'est la forme linéarisée par transformation des variables, pour permettre l'estimation des paramètres a et b par régression simple, du modèle :

$$V = A \times H^b \quad - (2)$$

où $A = 10^a$; sous cette forme, on peut calculer des points de la courbe ajustée pour les reporter sur papier millimétré simple.

On calcule les sommes, les sommes de carrés et de produits croisés des données transformées (tableau A21) d'une part pour chacun des jeux de données séparément, d'autre part pour l'ensemble des deux jeux :

Statistique	Origine des données		
	I	II	E(ensemble)
Σy	18,38954	20,28547	38,67501
Σy^2	35,37950	41,39075	76,77025
Σx	12,44696	12,65166	25,09862
Σx^2	15,67703	16,05102	31,72805
Σxy	23,43337	25,76286	49,18623
n	10	10	20
$SCy = \Sigma y^2 - (\Sigma y)^2/n$	1,56198	0,24072	1,98243
$SCx = \Sigma x^2 - (\Sigma x)^2/n$	0,18435	0,04457	0,23101
$SPxy = \Sigma xy - \Sigma x \times \Sigma y/n$	0,53398	0,09837	0,65176
<i>Coefficients :</i>			
$b = SPxy/SCx$	2,89657	2,20716	2,82135
$a = (\Sigma y - b \times \Sigma x)/n$	- 1,76640	- 0,76388	- 1,60685
<i>Somme des carrés expliquée par la régression :</i>			
$SCE = b \times SPxy$	1,54671	0,21712	1,83884

A partir de ces calculs de base, on peut procéder à l'analyse de variance. On calcule pour cela les quantités suivantes, où Σ signifie somme de I et II :

(i) Somme des carrés "entre coefficients b" :

$$SCb = \Sigma SCE - (\Sigma SPxy)^2 / \Sigma SCx$$

soit pour l'exemple choisi :

$$\begin{aligned} SCb &= (1,54671 + 0,21712) - (0,53398 + 0,09837)^2 / (0,18435 + 0,04457) \\ &= 0,01708 \end{aligned}$$

(ii) Somme des carrés "entre coefficients a" :

$$\begin{aligned} SCa &= \Sigma SCy - \Sigma SCE \\ &= (1,56198 + 0,24072) - (1,54671 + 0,21712) \\ &= 0,03887 \end{aligned}$$

(iii) Somme des carrés résiduelle :

$$SCR = SCy_E - (SCa + SCb + SCE_E)$$

où l'indice E est la marque de statistiques concernant la régression d'ensemble, construite sur l'ensemble des données. Pour l'exemple choisi, on obtient :

$$\begin{aligned} SCR &= 1,98243 - (0,03887 + 0,01708 + 1,83884) \\ &= 0,08764 \end{aligned}$$

(iv) Si r est le nombre de régressions comparées et n_E le nombre total d'observations pour l'ensemble des régressions, voici les nombres de degrés de liberté des quantités ci-dessus :

Statistique	Nombre de d.d.l.
SCy_E	$n_E - 1$
SCR	$n_E - 2r$
SCb	$r - 1$
Sca	$r - 1$
SCE_E	1

On peut alors dresser le tableau d'analyse de variance pour la comparaison des régressions :

Source de variation	Somme de carrés	d.d.l.	Carré moyen	F-observé
Expliquée par la régression d'ensemble (SCE_E)	1,83884	1	1,83884	335,7 ***
Entre pentes (SCb)	0,01708	1	0,01708	3,1
Entre constantes (Sca)	0,03887	1	0,03887	7,1 *
Résiduelle (SCR)	0,08764	16	0,005478	-
Totale (SCy_E)	1,98243	19	-	-

Les valeurs de F-observé s'obtiennent en divisant le carré moyen concerné par le carré moyen résiduel $SCR/(n_E-2r)$. Les astérisques caractérisent le degré de significativité des diverses valeurs de F-observé : on le détermine en se référant aux tables statistiques de la variable de FISHER (ici à 1 et 16 d.d.l. pour les trois F-observés) pour les seuils de probabilité de 5% (*), 1% (**) et 0,1% (***). D'après le tableau ci-dessus, la régression d'ensemble est hautement significative (il y a une probabilité très faible d'obtenir cette valeur de F-observé en cas d'absence de liaison statistique). La différence entre les pentes des deux régressions n'est pas significative, c'est-à-dire qu'elle peut être due aux effets de l'échantillonnage au hasard sur les deux jeux de données. En revanche, les constantes présentent une différence significative au seuil de 5%, ce qui suggère que les régressions sont distinctes bien qu'ayant la même pente. On en conclut donc que l'on ne peut pas rassembler les données des deux plantations, pour faire des prévisions à partir d'une régression d'ensemble unique, sans perte de précision sur les résultats.

Cette méthode de comparaison de régressions, peut être appliquée à n'importe quel nombre de régressions simples ; il suffit, dans les calculs de SCA et SCB ci-dessus, de faire les sommes Σ sur toutes les régressions à comparer. DAWKINS (voir la bibliographie) propose une fiche de calcul permettant de comparer jusqu'à 4 régressions.

Utilisation de variables conditionnelles pour la comparaison d'équations de régression

Une variable conditionnelle ⁽¹⁾ est une variable qui prend pour chaque observation la valeur 0 ou 1 selon que cette observation fait partie ou non d'un sous-ensemble déterminé. On peut utiliser de telles variables pour comparer des équations de régression et pour l'ajustement simultané à plusieurs jeux de données (régression "sériee", voir § A29).

Reprenons l'exemple ci-dessus où l'on comparait les courbes volume-hauteur obtenues par régression simple sur deux plantations différentes, I et II. Pour chaque plantation, on a un modèle :

$$\log V = a + b \times \log H \quad (1)$$

On peut construire un modèle unique pour les deux jeux de données en introduisant une variable supplémentaire Z qui vaut 0 pour les données provenant de la plantation I, et 1 pour les données de la plantation II. Ce modèle est le suivant :

$$\log V = (a_1 + a_2 Z) + (b_1 + b_2 Z) \times \log H \quad (3)$$

Pour les observations où Z=0 (plantation I), il conduit à :

$$\log V = a_1 + b_1 \times \log H \quad (4)$$

mais quand Z=1 (plantation II), il devient :

$$\log V = (a_1 + a_2) + (b_1 + b_2) \times \log H \quad (5)$$

Pour ajuster l'équation (3) par régression linéaire, on ôte tout simplement les parenthèses et on obtient :

$$\log V = a_1 + a_2 \times Z + b_1 \times \log H + b_2 \times Z \log H \quad (6)$$

qui est un modèle de régression linéaire à 3 variables explicatives :

$$Y = a_1 + a_2 X_1 + b_1 X_2 + b_2 X_3$$

(¹) en anglais = conditional variable

Un programme de régression multiple donne, pour les coefficients et leurs t-observés, les valeurs suivantes :

Coefficient	valeur	t-observé
a_1	- 1,76640	- 12,3
a_2	1,00179	3,1 (**)
b_1	2,89657	25,2
b_2	- 0,68884	- 2,7 (*)

Ces valeurs de t-observé (à 16 degrés de liberté) conduisent à des conclusions un peu différentes de l'analyse de covariance. Il faut en effet examiner la significativité de a_2 et b_2 qui représentent respectivement les différences entre constantes (a_2) et entre pentes (b_2) des équations des deux plantations ; les deux sont très significativement différents de zéro au seuil de 5 % : on ne peut donc admettre ni l'égalité des pentes, ni celle des constantes.

Rappelons que l'analyse de covariance conduisait à rejeter l'égalité des constantes, mais à admettre celle des pentes. Cette différence dans les conclusions auxquelles conduisent les deux méthodes ne doit pas choquer, car les modèles statistiques et les hypothèses testées n'y sont pas les mêmes ; l'analyse de covariance pose la question :

"Est-ce que l'une ou l'autre des régressions (ou les deux) diffère significativement de la régression d'ensemble ?".

tandis que la technique utilisant des variables conditionnelles demande :

"Est-ce que la constante (a_1+a_2) ou la pente (b_1+b_2) de la seconde régression diffère de la constante (a_1) ou de la pente (b_1) de la première ?".

On notera que ce second test est plus sensible aux différences que le premier, mais qu'il est moins pertinent quand il s'agit de savoir si l'on peut sans risque regrouper les régressions.

Les valeurs des coefficients de chacune des deux régressions se déduisent immédiatement des résultats ci-dessus, et sont d'ailleurs identiques à celles que fournit l'analyse de covariance :

Plantation I : Constante = a_1 = - 1,76640
Pente = b_1 = 2,89657

Plantation II : Constante = a_1+a_2 = - 0,76461
Pente = b_1+b_2 = 2,20773

Si ces chiffres peuvent différer à partir de la troisième décimale de ceux qui avaient été obtenus, c'est que ces derniers avaient été calculés à la main avec seulement 5 à 8 chiffres significatifs alors que la régression avec variables conditionnelles a été traitée par ordinateur avec 15 chiffres significatifs.

L'utilisation de variables conditionnelles peut devenir assez complexe, mais elle présente toujours l'intérêt d'une grande souplesse. Par exemple, étant donné une variable expliquée Y et une variable explicative X, il suffit d'une variable conditionnelle Z pour construire un modèle à pente unique mais à constantes différentes pour deux jeux de données :

$$Y = C_0 + C_1Z + C_2X$$

Pour trois jeux de données, deux variables conditionnelles deviennent nécessaires, par exemple :

$$Z_1 = 0 \text{ pour les données du jeu 1} \\ 1 \text{ pour les données des jeux 2 et 3}$$

$$Z_2 = 0 \text{ pour les données des jeux 1 et 2} \\ 1 \text{ pour les données du jeu 3}$$

et le modèle permettant de comparer les régressions simples avec pentes et constantes différentes est alors :

$$Y = C_0 + C_1Z_1 + C_2Z_2 + C_3X + C_4Z_1X + C_5Z_2X$$

On trouvera d'autres informations sur l'utilisation des variables conditionnelles dans le paragraphe suivant.

A29 REGRESSION SERIEE (1) AVEC EMPLOI DE VARIABLES CONDITIONNELLES

Nous avons déjà parlé des variables conditionnelles, qui peuvent prendre la valeur 0 ou 1, dans le paragraphe précédent. Outre leur utilité dans la comparaison des régressions, elles peuvent servir à ajuster un modèle de régression à des données "sériées" (2).

On a des données sériées quand les mesures sont groupées par unités ou placettes d'échantillonnage, auquel cas la régression "intra-placette" peut différer de la régression d'ensemble. Le cas se présente, dans les études de production forestière, quand les mesures sont faites sur des placettes permanentes : les données sont alors groupées par placettes. En voici un exemple typique :

(1) en anglais = nested regression

(2) en anglais = nested data

Exemple : la figure A24 représente graphiquement des données hauteur-âge provenant de 5 placettes permanentes mesurées 3 ou 4 fois. Le problème est d'évaluer la "pente" moyenne de la relation hauteur-âge en utilisant un modèle parabolique de la forme :

$$H = b_0 + b_1A + b_2A^2 \quad (1)$$

où b_1 et b_2 sont supposés identiques pour toutes les placettes mais où b_0 peut varier pour exprimer un effet de la station.

Si on analyse les données de la figure A24 par régression en utilisant comme modèle l'équation (1) et en considérant toutes les observations comme distinctes et indépendantes les unes des autres, on obtient pour les coefficients les valeurs suivantes :

$$b_0 = 7,807$$

$$b_1 = 0,6939$$

$$b_2 = -0,008652$$

avec un coefficient de corrélation $R = 0,74$ et un F -observé = 7,76, avec 2 et 13 degrés de liberté. La régression est donc significative au seuil de 1%.

On peut voir sur la figure A24 que cette régression (en traits interrompus) sous-estime manifestement la pente moyenne des mesures.

Pour y remédier, il convient d'ajuster un modèle sérié, dans lequel on introduit les 4 variables artificielles P_2 à P_5 définies comme suit :

$P_2 = 1$ pour les données de la placette 2, 0 pour les autres données

$P_3 = 1$ pour les données de la placette 3, 0 pour les autres données

$P_4 = 1$ pour les données de la placette 4, 0 pour les autres données

$P_5 = 1$ pour les données de la placette 5, 0 pour les autres données.

Le modèle est le suivant :

$$H = a_1 + a_2P_2 + a_3P_3 + a_4P_4 + a_5P_5 + b_1A + b_2A^2 \quad (2)$$

où les a_i et b_i sont les coefficients de la régression. C'est toujours un modèle parabolique, mais avec une constante différente pour chaque placette. Les valeurs de cette constante sont :

Placette	Constante
1	a_1
2	a_1+a_2
3	a_1+a_3
4	a_1+a_4
5	a_1+a_5

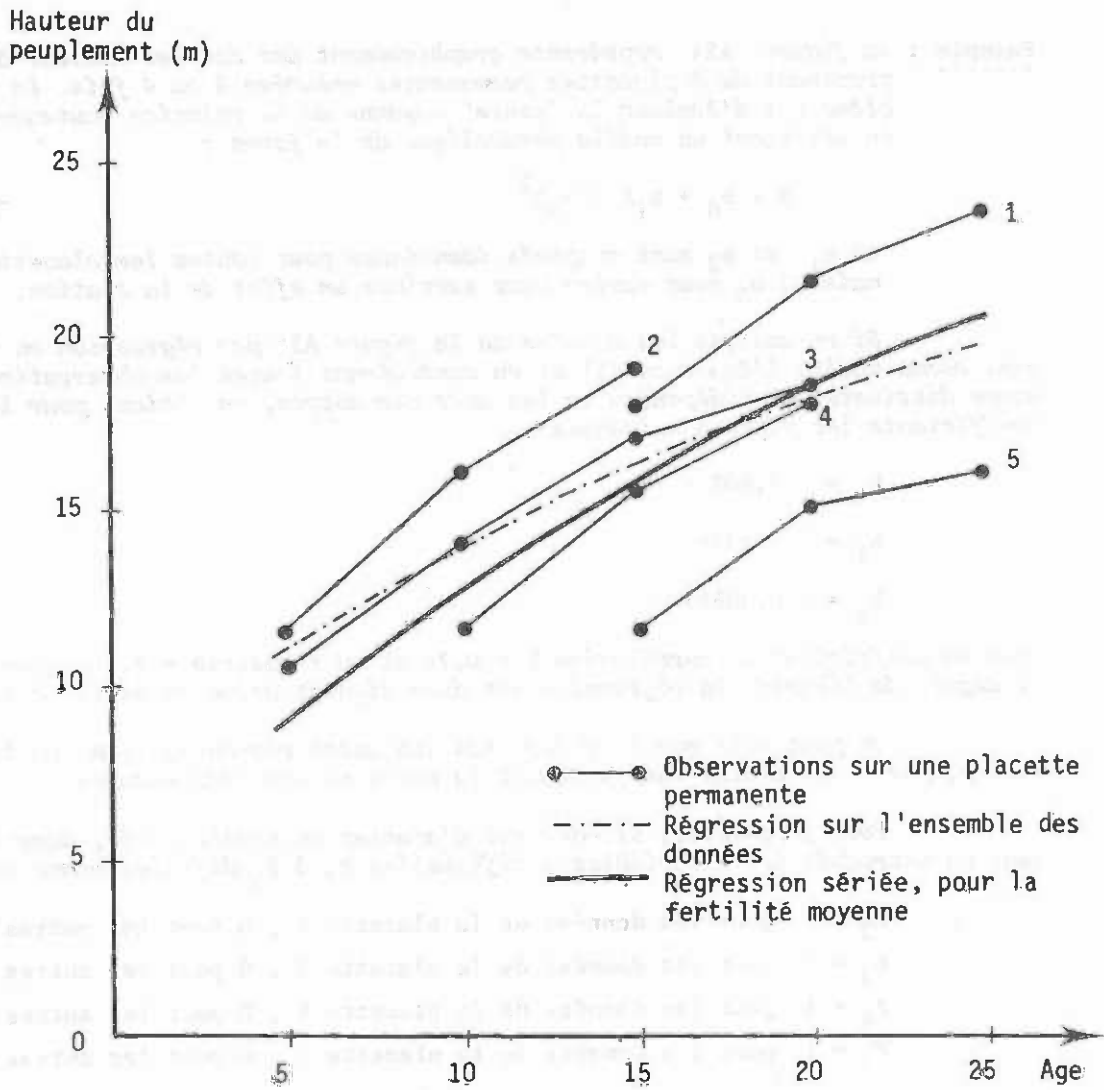


Figure A24 - Mesures hauteur-âge sur des placettes permanentes et comparaison des résultats obtenus par régression ordinaire et par régression sériée

Quand on ajuste le modèle (2) aux données de la figure A24, on obtient les estimations suivantes des coefficients :

a_1	7,432
a_2	0,3915
a_3	- 1,755
a_4	- 3,427
a_5	- 6,833
b_1	0,9192
b_2	- 0,01157

avec un coefficient de corrélation multiple de 0,99 et un F-observé de 70,88 à 6 et 9 degrés de liberté, largement significatif au seuil de 0,0001%. On obtient ainsi un faisceau de courbes parallèles dont l'allure est déterminée par b_1 et b_2 ; la figure A24 montre cette allure (en trait gras) pour une valeur de la constante égale à la valeur moyenne des 5 constantes. On peut voir que l'on obtient ainsi une bien meilleure image de l'évolution moyenne des placettes qu'en ajustant une régression à l'ensemble des données sans distinction.

La même technique peut évidemment être utilisée pour avoir des constantes et des pentes différentes sur chaque placette, ou encore une constante unique et des pentes différentes. Cette dernière possibilité peut s'avérer très utile avec les modèles qui donnent des courbes asymptotiques, comme l'équation de SCHUMACHER.

Ainsi, avec une variable expliquée Y et une seule variable explicative X, observées par exemple sur 3 placettes à plusieurs reprises, on peut ajuster les modèles suivants :

Pente unique (1), constantes différentes :

$$Y = a_1 + a_2P_2 + a_3P_3 + b_1X$$

Pentes et constantes différentes :

$$Y = a_1 + a_2P_2 + a_3P_3 + b_1X + b_2P_2X + b_3P_3X$$

Pentes différentes, constante unique (2) :

$$Y = a_1 + b_1X + b_2P_2X + b_3P_3X$$

où les a_i et b_i sont les coefficients de régression à estimer, X et Y des variables quelconques et les P_j des variables qui valent 1 pour la placette j et 0 pour les autres placettes.

(1) en anglais = common slope

(2) en anglais = common intercept

On peut étendre ces principes aux modèles à plusieurs variables explicatives, bien que le nombre de coefficients devienne vite grand. Il est possible de calculer à la main ces régressions sériées car, même si le nombre de variables est élevé, le caractère conditionnel (valeur 0 ou 1) de la plupart d'entre elles autorise de nombreux raccourcis dans les calculs. Dans ce cas, la méthode d'élimination de GAUSS-JORDAN, utilisée au § A25 pour inverser la matrice de covariance, n'est pas la plus rapide ni la plus pratique ; SEBER (1977, voir bibliographie) propose à ceux qui sont familiarisés avec l'algèbre matricielle des méthodes plus efficaces. En revanche, si l'on utilise un ordinateur, l'inefficacité relative des techniques de calcul classiques importe peu puisque le temps de traitement est de toute façon très court.

A3 RESOLUTION DES EQUATIONS

A31 RESOLUTION DE L'EQUATION DU 2ème DEGRE

On utilise très fréquemment l'équation du 2ème degré (parabole) comme modèle de régression à ajuster à des données qui suggèrent une faible courbure dans la relation entre x et y . Quand on a ajusté une telle équation pour estimer y en fonction de x , on a parfois besoin de déterminer x pour une valeur donnée de y ; on utilise alors la formule bien connue donnant les racines d'une équation du 2ème degré. Si le modèle de régression est :

$$y = b_0 + b_1x + b_2x^2$$

on l'écrit sous la forme :

$$ax^2 + bx + c = 0$$

où $a = b_2$, $b = b_1$ et $c = b_0 - y$; x est alors donné par :

$$x = (-b \pm \sqrt{b^2 - 4ac}) / 2a$$

On aura normalement deux solutions, selon que l'on donne au signe placé devant le radical la valeur + ou - . Dans un problème donné, il sera en général évident qu'une seule de ces deux solutions aura un sens : si la variable x est un diamètre et si les deux solutions obtenues sont +20 et -15 il est évident que seule la solution positive a un sens.

Quand la quantité sous le radical, $b^2 - 4ac$, est négative, l'équation n'a pas de solution réelle. Ceci s'explique généralement, en foresterie, par une erreur dans l'estimation des coefficients ou par une valeur de y trop grande ou trop petite. Supposons par exemple une fonction hauteur-âge dont la valeur du maximum est 42 m : si on cherche à résoudre cette équation pour voir à quel âge le peuplement atteindra 50 m, on se heurtera à une impossibilité car il n'y a pas de solution.

A32 RESOLUTION GRAPHIQUE DES EQUATIONS

La plupart des modèles complexes conduisent à des équations qu'on ne peut pas résoudre algébriquement : on est alors conduit à employer une méthode graphique ou numérique. La méthode de résolution graphique est à peu près évidente : elle consiste à tracer point par point la fonction :

$$y = f(x)$$

en calculant y pour des valeurs déterminées de x ; on relie ensuite les points obtenus par une courbe régulière ⁽¹⁾ puis on détermine la valeur de x correspondant à la valeur de y pour laquelle on cherche la solution.

(1) en anglais = smooth curve

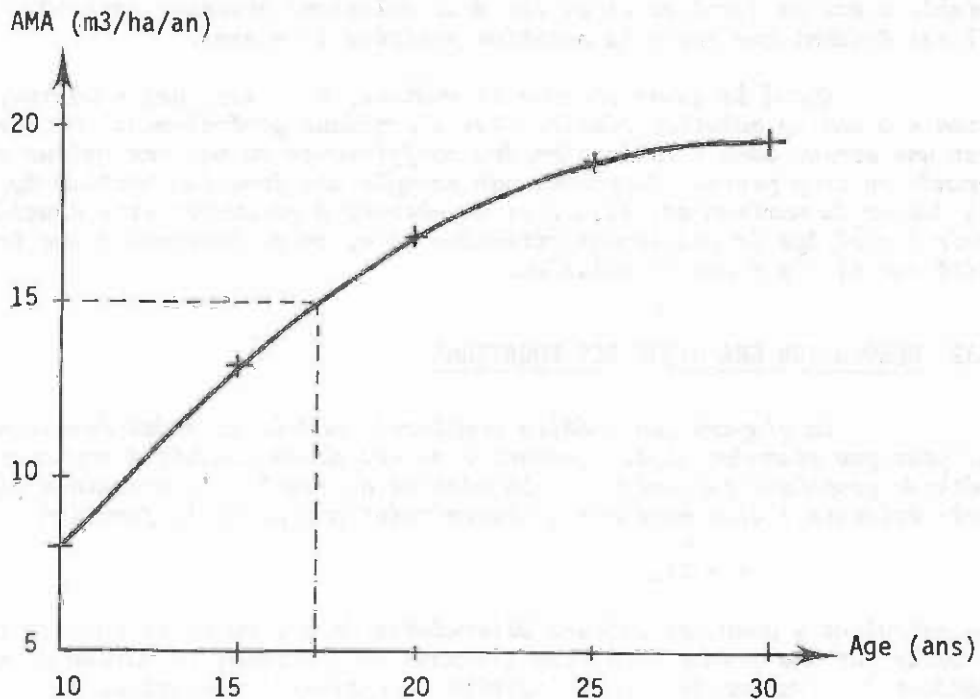
Supposons par exemple que l'on ait ajusté l'équation :

$$\log \text{AMA} = - 0,8892 - 0,03055 A + 2,097 \log A \quad (1)$$

reliant l'accroissement moyen annuel en volume (AMA) d'une plantation à son âge (A), et que l'on veuille la résoudre pour déterminer l'âge auquel cet AMA atteint pour la première fois 15 m³/ha/an. Pour résoudre graphiquement ce problème, on commence par calculer, à l'aide de l'équation (1), l'AMA pour une série d'âges à l'intérieur de laquelle on peut espérer que se trouve la réponse cherchée ; le tableau ci-dessous donne la valeur de l'AMA pour des âges allant de 10 à 30 ans :

Age	AMA
10	7,99
15	13,15
20	16,91
25	18,99
30	19,58

On porte sur un graphique Age-AMA les points correspondants (marqués d'un + sur le graphique ci-après) et on les relie à la main par une courbe régulière. On peut alors lire sur ce graphique la valeur de l'abscisse correspondant à la valeur connue de l'ordonnée, ce qui donne dans notre exemple un âge de 17,2 ans pour un accroissement moyen annuel de 15 m³/ha/an.



Les principaux inconvénients de la méthode de résolution graphique sont les suivants :

- (i) la précision est de l'ordre de 0,5% de l'étendue de l'échelle des x
- (ii) c'est une méthode lente, que l'on ne peut appliquer qu'à la main ; on ne peut pas l'utiliser dans un programme d'ordinateur, ni la programmer sur une calculatrice.

A33 RESOLUTION NUMERIQUE DES EQUATIONS

Il existe plusieurs méthodes numériques de résolution des équations, mais l'une des plus simples et des plus sûres est celle que l'on appelle "méthode dichotomique" ⁽¹⁾. Les étapes en sont les suivantes :

- (i) Définir l'équation à résoudre sous la forme :

$$y = f(x)$$

où y est une valeur connue et x la valeur inconnue à déterminer par cette méthode. Il est nécessaire de savoir a priori que la solution recherchée, x, appartient à un intervalle déterminé $[x_a, x_b]$ tel que :

- [1] la fonction f(x) soit continue sur $[x_a, x_b]$
- [2] $f(x_a) - y$ et $f(x_b) - y$ soient de signes contraires.

On appelle ϵ ($= +1$ ou -1) le signe de $f(x_b) - f(x_a)$

- (ii) Calculer la valeur x_1 correspondant au milieu de $[x_a, x_b]$

$$x_1 = (x_a + x_b) / 2$$

- (iii) Calculer la valeur y_1 que prend la fonction pour x_1 :

$$y_1 = f(x_1)$$

- (iv) Si : $\epsilon(y_1 - y) > 0$, alors remplacer x_b par x_1 ;

$\epsilon(y_1 - y) < 0$, alors remplacer x_a par x_1 ;

$y_1 - y = 0$, alors la solution est x_1 ; il y a peu de chance de rencontrer ce dernier cas dans la pratique.

- (v) Reprendre les étapes (ii) à (iv) et les répéter n fois, n étant déterminé par la précision que l'on souhaite obtenir sur la solution recherchée. Cette précision est donnée, en fonction des valeurs initiales x_a, x_b et du nombre d'itérations n, par la formule :

⁽¹⁾ en anglais = bisection method

$$\text{erreur maximum} = (x_b - x_a) / 2^n$$

Après n itérations, la dernière valeur de x_i peut être prise comme solution de l'équation, avec la précision définie ci-dessus. En fait, la détermination fine d'une valeur optimale de n dans chaque cas n'est pas cruciale, vu la rapidité des ordinateurs : on peut donc choisir a priori une valeur de n qui assure une précision confortable pour toute une catégorie de problèmes. On a en effet :

Nombre d'itérations	Précision en % de la longueur de l'intervalle $[x_a, x_b]$
5	3 %
8	0,3 %
10	0,1 %
15	0,003 %
20	0,0001 %
30	0,0000001 %

Dans la plupart des problèmes forestiers, il suffit de 15 itérations pour obtenir une précision correcte.

Pour prendre un exemple d'application de la méthode dichotomique, supposons que l'on ait ajusté, comme au paragraphe précédent, une équation reliant l'accroissement moyen annuel (AMA) à l'âge (A) :

$$\log \text{AMA} = - 0,8892 - 0,03055 A + 2,097 \log A$$

et que l'on souhaite la résoudre pour déterminer l'âge auquel l'AMA atteint 15m³/ha/an. En prenant $x_a = 1$ an, $x_b = 30$ ans, $y = 15$ m³/ha/an et en répétant les étapes (ii) à (iv) définies ci-dessus, on obtient le tableau suivant :

i	x_a	x_b	x_i	y_i
1	1,000	30,000	15,500	13,595
2	15,500	30,000	22,750	18,254
3	15,500	22,750	19,125	16,370
4	15,500	19,125	17,312	15,091
5	15,500	17,312	16,406	14,370
6	16,406	17,312	16,859	14,737
7	16,859	17,312	17,085	14,916
8	17,085	17,312	17,199	15,004
9	17,085	17,199	17,142	14,960
10	17,142	17,199	17,170	14,982
11	17,170	17,199	17,185	14,993
12	17,185	17,199	17,192	14,993
13	17,192	17,199	17,195	15,001
14	17,192	17,195	17,193	15,000
15	17,192	17,193	17,193	14,999

Le principal inconvénient de cette méthode par rapport à d'autres méthodes de résolution numérique (notamment la méthode de NEWTON) est que la convergence, au fil des itérations, est relativement lente ; elle implique en outre que l'on soit capable de définir a priori un intervalle incluant la solution. Ces inconvénients sont rarement gênants en recherche forestière ; s'ils le devenaient, il faudrait se référer à un manuel d'analyse numérique, comme celui de STARK (1970) par exemple, pour avoir des conseils plus précis.

A4 AJUSTEMENT DES MODELES NON LINEAIRES

Un modèle non linéaire est un modèle que l'on ne peut pas transformer, par transformation des variables, en modèle linéaire : on ne peut donc pas en estimer directement les paramètres par régression linéaire (méthode des moindres carrés) comme au chapitre A2.

On peut, par exemple, transformer l'équation :

$$Y = a X^b$$

en prenant le logarithme des deux membres pour obtenir le modèle linéaire :

$$\log Y = \log a + b \log X$$

ce qui permet d'estimer directement $\log a$ et b par régression linéaire.

Mais l'équation :

$$Y = a + b X^c$$

dans laquelle a, b, c sont des coefficients à déterminer, ne peut pas être transformée directement en modèle linéaire : on dit qu'il s'agit d'un modèle, ou d'une équation, non linéaire.

Nous ne pouvons traiter, dans le cadre de ce manuel, que de l'ajustement des fonctions non linéaires les plus simples couramment utilisées dans les études d'accroissement et de production ; elles présentent les caractéristiques suivantes :

- (i) Il n'y a qu'une variable explicative (X) et une variable expliquée (Y).
- (ii) Il y a trois paramètres (que nous appellerons a, b , et k) à estimer par ajustement aux données.
- (iii) Si la valeur du paramètre k est supposée connue, on peut estimer a et b par régression linéaire simple en utilisant des transformations appropriées.

Parmi les équations qui satisfont à ces conditions, les suivantes sont couramment utilisées :

- 1 $Y = a + b X^k$
- 2 $Y = a e^{bX^k}$
- 3 $Y = a \left(1 - e^{-kX} \right)^b$

Il existe évidemment d'autres modèles non linéaires ajustables, mais nous n'en parlerons pas ici, car ils sont moins fréquemment utilisés en matière forestière.

Le problème principal est l'estimation du paramètre k . Il est facile pour une valeur connue de k , de transformer les trois équations ci-dessus en modèles linéaires :

1a) $Y = a + b (X^k)$

où la variable explicative est X^k

2a) $\log_e Y = \log_e a + b (X^k)$

où la variable explicative est encore X^k et la variable expliquée $\log_e Y$

3a) $\log_e Y = \log_e a + b \log_e (1 - e^{-kX})$

où la variable explicative est $\log_e (1 - e^{-kX})$ et la variable expliquée $\log_e Y$.

Pour une valeur de k fixée, on peut ajuster ces trois modèles "linéarisés" par régression simple. Nous allons voir que l'on peut déterminer graphiquement la valeur de k qui fournit le meilleur ajustement.

Le principe en est simple : on calcule, par la méthode classique décrite au paragraphe A21, la somme des carrés résiduelle du modèle linéarisé pour une série de valeurs de k ; on porte sur un graphique ces sommes de carrés résiduelles en fonction de la valeur de k , et on trace à la main une courbe régulière passant par les points ainsi définis ; la valeur de k correspondant au minimum de cette courbe constitue la meilleure estimation de ce paramètre, et la régression linéaire faite avec cette valeur de k donne les meilleures estimations correspondantes des paramètres a et b .

Prenons comme exemple les données représentées sur la figure A41, qui sont des hauteurs observées sur une placette à différents âges ; on souhaite y ajuster le modèle :

$$Y = a (1 - e^{-kX})^b$$

Y étant la hauteur et X l'âge du peuplement. La forme linéarisée du modèle

$$\log Y = \log a + b \log (1 - e^{-kX})$$

permet d'estimer la constante $\log a$ et la pente b de la régression linéaire

simple de $\log Y$ sur $\log (1 - e^{-kX})$ pour n'importe quelle valeur de k . Pour le moment, c'est surtout la somme des carrés résiduelle SCR qui nous intéresse dans ces régressions (voir l'imprimé A2 au paragraphe A21). Avec les données de la figure A41, on obtient pour une série de valeurs de k les sommes de carrés résiduelles ci-dessous :

k	somme des carrés résiduelle
0,05	0,2828
0,15	0,0302
0,25	0,0732
0,35	0,2657
0,45	0,4887

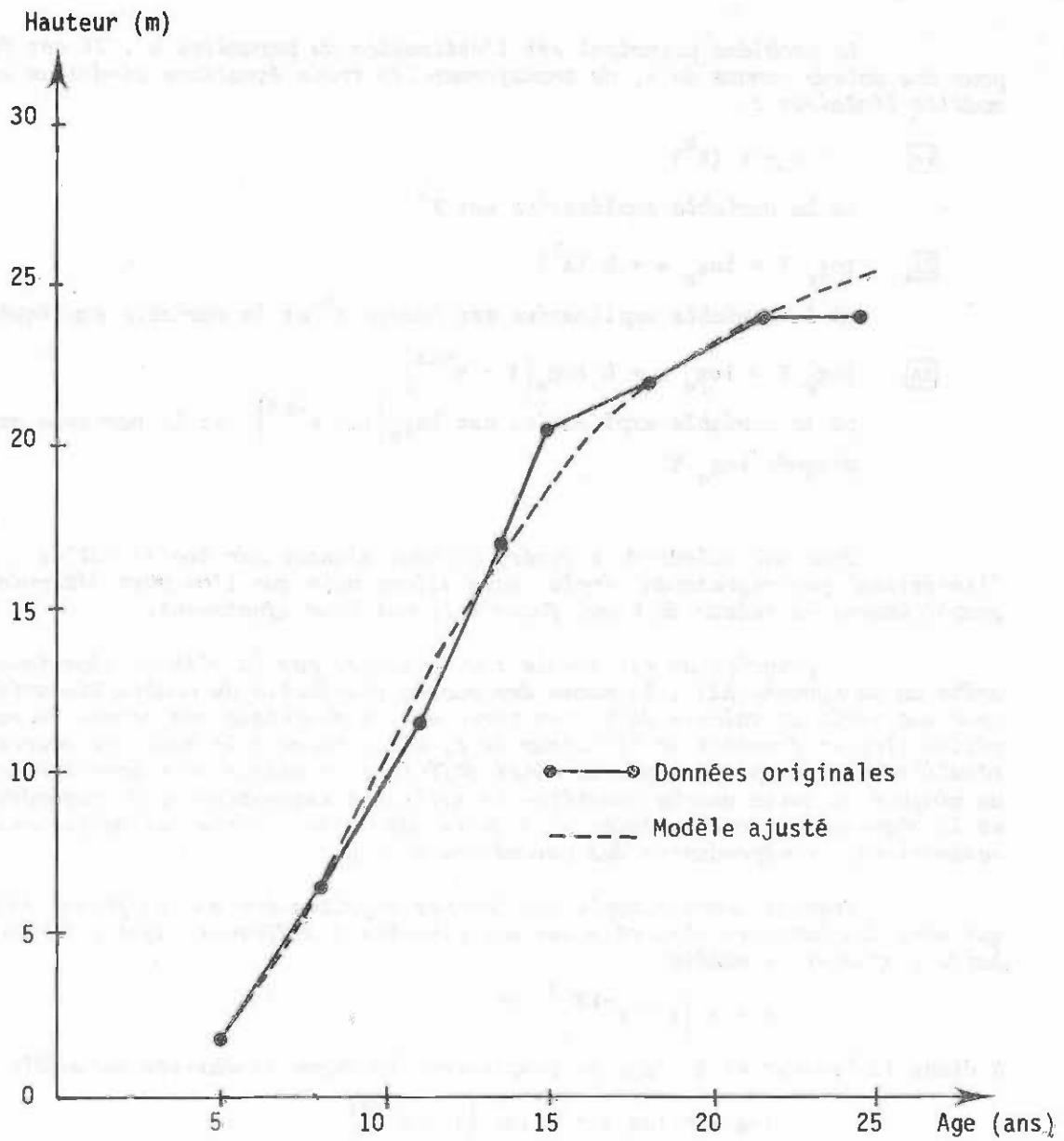


Figure A41 - Données hauteur-âge observées sur une placette et un modèle correspondant (en trait interrompu) ajusté par la méthode de régression non linéaire exposée dans le texte.

Tableau des données observées	
Age (ans)	Hauteur (m)
5	1.5
8	6.5
11	11.5
14	17.0
15	20.5
18	21.5
22	23.5
25	23.5

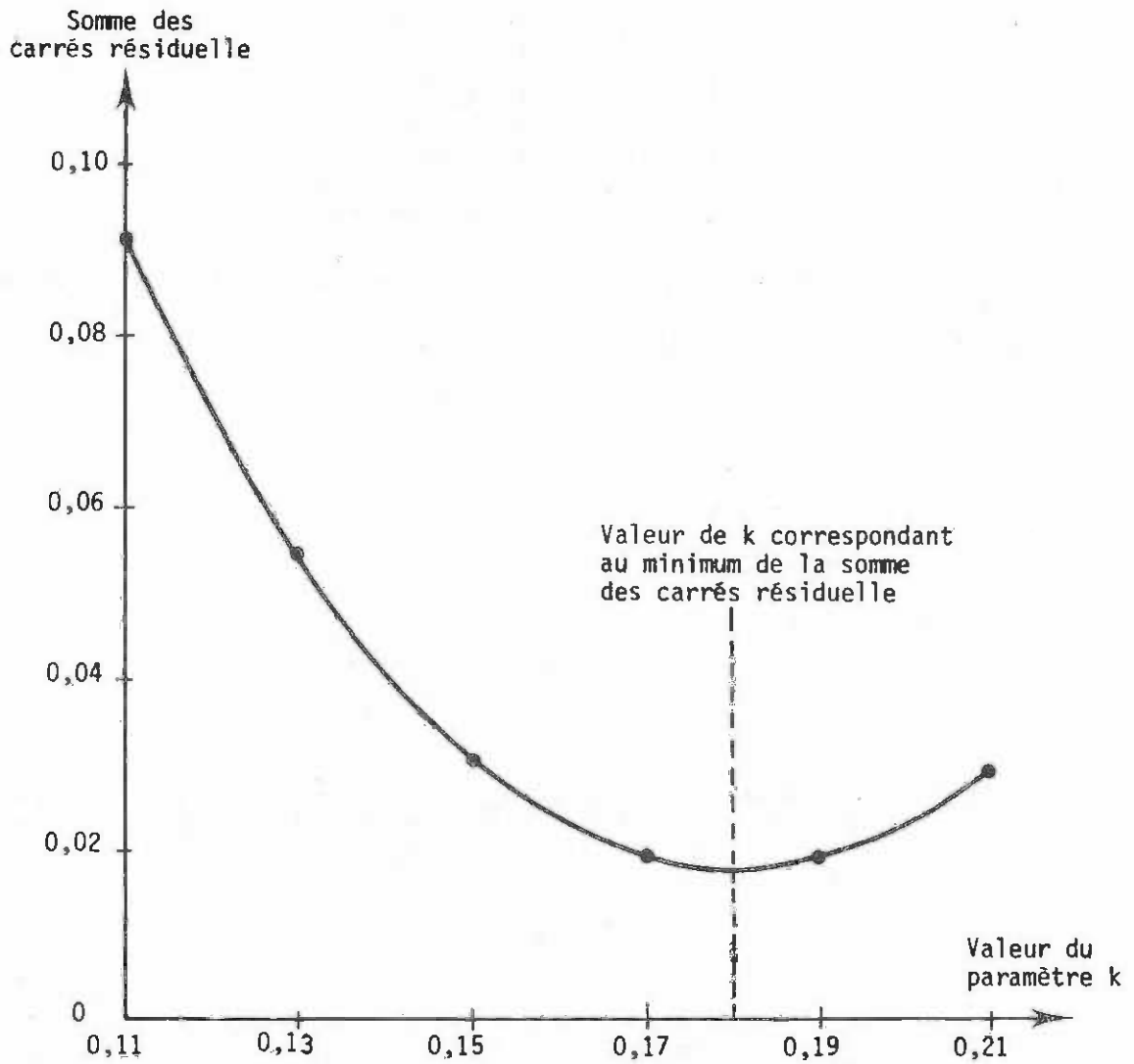


Figure A42 - Détermination graphique de la valeur du paramètre k qui minimise la somme des carrés résiduelle dans une régression non linéaire.

Il semble que le meilleur ajustement soit obtenu en prenant pour k une valeur voisine de 0,15. En explorant plus finement autour de cette valeur, on obtient :

k	SCR
0,11	0,0916
0,13	0,0543
0,15	0,0302
0,17	0,0187
0,19	0,0187
0,21	0,0287
0,23	0,0474

Si on porte ces valeurs sur un graphique (figure A42), on constate que le minimum est très proche de $k = 0,18$; si on calcule les valeurs des coefficients de la régression pour cette valeur de k , on trouve :

$$\log_e a = 3,290 \quad \text{soit } a = 26,83$$
$$b = 5,199$$

et l'équation correspondante est :

$$H = 26,83 \left(1 - e^{-0,18A} \right)^{5,199}$$

Elle est représentée en trait interrompu sur la figure A41.

On peut aussi étendre cette approche dans deux directions que nous nous contenterons de présenter rapidement car leur exposé détaillé sortirait du cadre de ce manuel.

- (i) Tout modèle contenant un seul paramètre non linéaire (et donc ajustable par régression linéaire quand la valeur de ce paramètre est fixée) peut être ajusté d'une manière analogue. Par exemple, avec le modèle :

$$Y = b_0 + b_1 X_1^k + b_2 X_2 + b_3 X_1^{(1-k)} X_2$$

on peut estimer les coefficients b_i par régression linéaire multiple pour une série de valeurs de k , porter sur un graphique les sommes de carrés résiduelles obtenues en fonction de la valeur de k , déterminer la valeur de k correspondant à la SCR minimum et estimer enfin les coefficients b_i pour cette valeur de k .

- (ii) La recherche de la valeur de k qui minimise la somme des carrés résiduelle peut être programmée, si bien que l'estimation de tous les paramètres (y compris le paramètre non linéaire) peut être faite par un programme d'ordinateur. C'est très facile à réaliser quand le modèle ne comporte qu'un paramètre non linéaire : on peut même obtenir des programmes assez courts pour certaines calculatrices programmables. On peut trouver dans ROYCE-SADLER (1975) des méthodes de minimisation adaptées.

Nous avons exposé cette méthode pour les chercheurs qui ne disposent pas d'un ordinateur conséquent. Si l'on a des moyens de calcul importants, il vaut mieux employer l'une des variantes de la méthode de GAUSS-NEWTON présentées dans les manuels classiques comme DRAPER and SMITH (1966). Ces méthodes étant difficiles à comprendre et à mettre en oeuvre, il est préférable de chercher à se procurer un programme tout fait, compatible avec l'ordinateur dont on dispose. Il en existe beaucoup ; on peut signaler en particulier le programme "SNIFTA" (Small nonlinear fitting algorithm), écrit en FORTRAN ANSI 66 et disponible auprès de :

Biometrics Section
Unit of tropical silviculture
Commonwealth Forestry Institute
South Parks Road, OXFORD OX1 3RB
United Kingdom.

Il faut noter que les fonctions courantes sont définies dans le chapitre 2. Les fonctions courantes sont définies dans le chapitre 2. Les fonctions courantes sont définies dans le chapitre 2.

Unité de travail
Bureau de l'Institut
1000

A N N E X E B

TABLES DES FONCTIONS COURANTES UTILISEES DANS
LES MODELES DECRITS DANS LE TEXTE

Table 1 : Logarithme népérien (base e) de X, pour X variant de 1 à 105

X	ln X	X	ln X	X	ln X
1	0.00000	36	3.58351	71	4.26267
2	0.69314	37	3.61091	72	4.27666
3	1.09861	38	3.63758	73	4.29045
4	1.38629	39	3.66356	74	4.30406
5	1.60943	40	3.68887	75	4.31748
6	1.79175	41	3.71357	76	4.33073
7	1.94591	42	3.73766	77	4.34380
8	2.07944	43	3.76120	78	4.35670
9	2.19722	44	3.78418	79	4.36944
10	2.30258	45	3.80666	80	4.38202
11	2.39789	46	3.82864	81	4.39444
12	2.48490	47	3.85014	82	4.40671
13	2.56494	48	3.87120	83	4.41884
14	2.63905	49	3.89182	84	4.43081
15	2.70805	50	3.91202	85	4.44265
16	2.77258	51	3.93182	86	4.45434
17	2.83321	52	3.95124	87	4.46590
18	2.89037	53	3.97029	88	4.47733
19	2.94443	54	3.98898	89	4.48863
20	2.99573	55	4.00733	90	4.49980
21	3.04452	56	4.02535	91	4.51085
22	3.09104	57	4.04305	92	4.52178
23	3.13549	58	4.06044	93	4.53259
24	3.17805	59	4.07753	94	4.54329
25	3.21887	60	4.09434	95	4.55387
26	3.25809	61	4.11087	96	4.56434
27	3.29583	62	4.12713	97	4.57471
28	3.33220	63	4.14313	98	4.58496
29	3.36729	64	4.15888	99	4.59511
30	3.40119	65	4.17438	100	4.60517
31	3.43398	66	4.18965	101	4.61512
32	3.46573	67	4.20469	102	4.62497
33	3.49650	68	4.21950	103	4.63472
34	3.52636	69	4.23410	104	4.64439
35	3.55534	70	4.24849	105	4.65396

Table 2 : Fonction exponentielle $\exp(X) = e^X$, pour X variant de 0 à 4.99

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09
0.1	1.10	1.11	1.12	1.13	1.15	1.16	1.17	1.18	1.19	1.20
0.2	1.22	1.23	1.24	1.25	1.27	1.28	1.29	1.30	1.32	1.33
0.3	1.34	1.36	1.37	1.39	1.40	1.41	1.43	1.44	1.46	1.47
0.4	1.49	1.50	1.52	1.53	1.55	1.56	1.58	1.59	1.61	1.63
0.5	1.64	1.66	1.68	1.69	1.71	1.73	1.75	1.76	1.78	1.80
0.6	1.82	1.84	1.85	1.87	1.89	1.91	1.93	1.95	1.97	1.99
0.7	2.01	2.03	2.05	2.07	2.09	2.11	2.13	2.15	2.18	2.20
0.8	2.22	2.24	2.27	2.29	2.31	2.33	2.36	2.38	2.41	2.43
0.9	2.45	2.48	2.50	2.53	2.55	2.58	2.61	2.63	2.66	2.69
1.0	2.71	2.74	2.77	2.80	2.82	2.85	2.88	2.91	2.94	2.97
1.1	3.00	3.03	3.06	3.09	3.12	3.15	3.18	3.22	3.25	3.28
1.2	3.32	3.35	3.38	3.42	3.45	3.49	3.52	3.56	3.59	3.63
1.3	3.66	3.70	3.74	3.78	3.81	3.85	3.89	3.93	3.97	4.01
1.4	4.05	4.09	4.13	4.17	4.22	4.26	4.30	4.34	4.39	4.43
1.5	4.48	4.52	4.57	4.61	4.66	4.71	4.75	4.80	4.85	4.90
1.6	4.95	5.00	5.05	5.10	5.15	5.20	5.25	5.31	5.36	5.41
1.7	5.47	5.52	5.58	5.64	5.69	5.75	5.81	5.87	5.92	5.98
1.8	6.04	6.11	6.17	6.23	6.29	6.35	6.42	6.48	6.55	6.61
1.9	6.68	6.75	6.82	6.88	6.95	7.02	7.09	7.17	7.24	7.31
2.0	7.38	7.46	7.53	7.61	7.69	7.76	7.84	7.92	8.00	8.08
2.1	8.16	8.24	8.33	8.41	8.49	8.58	8.67	8.75	8.84	8.93
2.2	9.02	9.11	9.20	9.29	9.39	9.48	9.58	9.67	9.77	9.87
2.3	9.97	10.07	10.17	10.27	10.38	10.48	10.59	10.69	10.80	10.91
2.4	11.02	11.13	11.24	11.35	11.47	11.58	11.70	11.82	11.94	12.06
2.5	12.18	12.30	12.42	12.55	12.67	12.80	12.93	13.06	13.19	13.32
2.6	13.46	13.59	13.73	13.87	14.01	14.15	14.29	14.43	14.58	14.73
2.7	14.87	15.02	15.18	15.33	15.48	15.64	15.79	15.95	16.11	16.28
2.8	16.44	16.60	16.77	16.94	17.11	17.28	17.46	17.63	17.81	17.99
2.9	18.17	18.35	18.54	18.72	18.91	19.10	19.29	19.49	19.68	19.88
3.0	20.08	20.28	20.49	20.69	20.90	21.11	21.32	21.54	21.75	21.97
3.1	22.19	22.42	22.64	22.87	23.10	23.33	23.57	23.80	24.04	24.28
3.2	24.53	24.77	25.02	25.27	25.53	25.79	26.04	26.31	26.57	26.84
3.3	27.11	27.38	27.66	27.93	28.21	28.50	28.78	29.07	29.37	29.66
3.4	29.96	30.26	30.56	30.87	31.18	31.50	31.81	32.13	32.45	32.78
3.5	33.11	33.44	33.78	34.12	34.46	34.81	35.16	35.51	35.87	36.23
3.6	36.59	36.96	37.33	37.71	38.09	38.47	38.86	39.25	39.64	40.04
3.7	40.44	40.85	41.26	41.67	42.09	42.52	42.94	43.38	43.81	44.25
3.8	44.70	45.15	45.60	46.06	46.52	46.99	47.46	47.94	48.42	48.91
3.9	49.40	49.89	50.40	50.90	51.41	51.93	52.45	52.98	53.51	54.05
4.0	54.59	55.14	55.70	56.26	56.82	57.39	57.97	58.55	59.14	59.73
4.1	60.34	60.94	61.55	62.17	62.80	63.43	64.07	64.71	65.36	66.02
4.2	66.68	67.35	68.03	68.71	69.40	70.10	70.80	71.52	72.24	72.96
4.3	73.69	74.44	75.18	75.94	76.70	77.47	78.25	79.04	79.83	80.64
4.4	81.45	82.26	83.09	83.93	84.77	85.62	86.48	87.35	88.23	89.12
4.5	90.01	90.92	91.83	92.75	93.69	94.63	95.58	96.54	97.51	98.49
4.6	99.48	100.48	101.49	102.51	103.54	104.58	105.63	106.69	107.77	108.85
4.7	109.94	111.05	112.16	113.29	114.43	115.58	116.74	117.91	119.10	120.30
4.8	121.51	122.73	123.96	125.21	126.46	127.74	129.02	130.32	131.63	132.95
4.9	134.28	135.63	137.00	138.37	139.77	141.17	142.59	144.02	145.47	146.93

Table 3 : Table de $1/X^\alpha$ pour X variant de 2 à 50 et α variant de 0.2 à 2

X	Valeur de l'exposant α									
	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
2	.87055	.75785	.65975	.57434	.50000	.43527	.37892	.32987	.28717	.25000
3	.80274	.64439	.51728	.41524	.33333	.26758	.21479	.17242	.13841	.11111
4	.75785	.57434	.43527	.32987	.25000	.18946	.14358	.10881	.08246	.06250
5	.72477	.52530	.38073	.27594	.20000	.14495	.10506	.07614	.05518	.04000
6	.69882	.48835	.34127	.23849	.16666	.11647	.08139	.05687	.03974	.02777
7	.67761	.45915	.31112	.21082	.14285	.09680	.06559	.04444	.03011	.02040
8	.65975	.43527	.28717	.18946	.12500	.08246	.05440	.03589	.02368	.01562
9	.64439	.41524	.26758	.17242	.11111	.07159	.04613	.02973	.01915	.01234
10	.63095	.39810	.25118	.15848	.10000	.06309	.03981	.02511	.01584	.01000
11	.61904	.38321	.23722	.14685	.09090	.05627	.03483	.02156	.01335	.00826
12	.60836	.37010	.22516	.13697	.08333	.05069	.03084	.01876	.01141	.00694
13	.59870	.35844	.21460	.12848	.07692	.04605	.02757	.01650	.00988	.00591
14	.58989	.34797	.20526	.12108	.07142	.04213	.02485	.01466	.00864	.00510
15	.58181	.33850	.19694	.11458	.06666	.03878	.02256	.01312	.00763	.00444
16	.57434	.32987	.18946	.10881	.06250	.03589	.02061	.01184	.00680	.00390
17	.56742	.32197	.18269	.10366	.05882	.03337	.01893	.01074	.00609	.00346
18	.56097	.31469	.17653	.09903	.05555	.03116	.01748	.00980	.00550	.00308
19	.55494	.30796	.17090	.09484	.05263	.02920	.01620	.00899	.00499	.00277
20	.54928	.30170	.16572	.09102	.05000	.02746	.01508	.00828	.00455	.00250
21	.54394	.29587	.16094	.08754	.04761	.02590	.01408	.00766	.00416	.00226
22	.53890	.29042	.15651	.08434	.04545	.02449	.01320	.00711	.00383	.00206
23	.53413	.28530	.15239	.08139	.04347	.02322	.01240	.00662	.00353	.00189
24	.52961	.28048	.14855	.07867	.04166	.02206	.01168	.00618	.00327	.00173
25	.52530	.27594	.14495	.07614	.04000	.02101	.01103	.00579	.00304	.00160
26	.52120	.27165	.14158	.07379	.03846	.02004	.01044	.00544	.00283	.00147
27	.51728	.26758	.13841	.07159	.03703	.01915	.00991	.00512	.00265	.00137
28	.51353	.26371	.13542	.06954	.03571	.01834	.00941	.00483	.00248	.00127
29	.50994	.26004	.13260	.06762	.03448	.01758	.00896	.00457	.00233	.00118
30	.50649	.25653	.12993	.06581	.03333	.01688	.00855	.00433	.00219	.00111
31	.50318	.25319	.12740	.06410	.03225	.01623	.00816	.00410	.00206	.00104
32	.50000	.25000	.12500	.06250	.03125	.01562	.00781	.00390	.00195	.00097
33	.49693	.24694	.12271	.06098	.03030	.01505	.00748	.00371	.00184	.00091
34	.49397	.24401	.12053	.05954	.02941	.01452	.00717	.00354	.00175	.00086
35	.49111	.24119	.11845	.05817	.02857	.01403	.00689	.00338	.00166	.00081
36	.48835	.23849	.11647	.05687	.02777	.01356	.00662	.00323	.00157	.00077
37	.48569	.23589	.11457	.05564	.02702	.01312	.00637	.00309	.00150	.00073
38	.48310	.23339	.11275	.05447	.02631	.01271	.00614	.00296	.00143	.00069
39	.48060	.23097	.11100	.05335	.02564	.01232	.00592	.00284	.00136	.00065
40	.47817	.22865	.10933	.05228	.02500	.01195	.00571	.00273	.00130	.00062
41	.47582	.22640	.10772	.05125	.02439	.01160	.00552	.00262	.00125	.00059
42	.47353	.22423	.10618	.05028	.02380	.01127	.00533	.00252	.00119	.00056
43	.47130	.22213	.10469	.04934	.02325	.01096	.00516	.00243	.00114	.00054
44	.46914	.22009	.10325	.04844	.02272	.01066	.00500	.00234	.00110	.00051
45	.46704	.21812	.10187	.04758	.02222	.01037	.00484	.00226	.00105	.00049
46	.46499	.21622	.10054	.04675	.02173	.01010	.00470	.00218	.00101	.00047
47	.46299	.21436	.09925	.04595	.02127	.00985	.00456	.00211	.00097	.00045
48	.46105	.21257	.09800	.04518	.02083	.00960	.00442	.00204	.00094	.00043
49	.45915	.21082	.09680	.04444	.02040	.00937	.00430	.00197	.00090	.00041
50	.45730	.20912	.09563	.04373	.02000	.00914	.00418	.00191	.00087	.00040

Table 4 : Espacement et facteur d'espacement en %, en fonction du nombre de tiges par hectare et de la hauteur du peuplement (plantation en carré).

Espacement Nbre/ha	2.0 2500	2.5 1600	3.0 1111	3.5 816	4.0 625	4.5 494	5.0 400	6.0 278	7.0 204	8.0 m. 156
Hauteur	----- Facteur d'espacement % -----									
5	40	50	60	70	80	89	100	119	140	160
6	33	41	50	58	66	74	83	99	116	133
7	28	35	42	50	57	64	71	85	100	114
8	25	31	37	43	50	56	62	74	87	100
9	22	27	33	38	44	49	55	66	77	88
10	20	25	30	35	40	44	50	59	70	80
11	18	22	27	31	36	40	45	54	63	72
12	16	20	25	29	33	37	41	49	58	66
13	15	19	23	26	30	34	38	46	53	61
14	14	17	21	25	28	32	35	42	50	57
15	13	16	20	23	26	29	33	39	46	53
16	12	15	18	21	25	28	31	37	43	50
17	11	14	17	20	23	26	29	35	41	47
18	11	13	16	19	22	24	27	33	38	44
19	10	13	15	18	21	23	26	31	36	42
20	10	12	15	17	20	22	25	29	35	40
21	9	11	14	16	19	21	23	28	33	38
22	9	11	13	15	18	20	22	27	31	36
23	8	10	13	15	17	19	21	26	30	34
24	8	10	12	14	16	18	20	24	29	33
25	8	10	12	14	16	17	20	23	28	32
26	7	9	11	13	15	17	19	23	26	30
27	7	9	11	12	14	16	18	22	25	29
28	7	8	10	12	14	16	17	21	25	28
29	6	8	10	12	13	15	17	20	24	27
30	6	8	10	11	13	14	16	19	23	26
31	6	8	9	11	12	14	16	19	22	25
32	6	7	9	10	12	14	15	18	21	25
33	6	7	9	10	12	13	15	18	21	24
34	5	7	8	10	11	13	14	17	20	23
35	5	7	8	10	11	12	14	17	20	22
36	5	6	8	9	11	12	13	16	19	22
37	5	6	8	9	10	12	13	16	18	21
38	5	6	7	9	10	11	13	15	18	21
39	5	6	7	8	10	11	12	15	17	20
40	5	6	7	8	10	11	12	14	17	20
41	4	6	7	8	9	10	12	14	17	19
42	4	5	7	8	9	10	11	14	16	19
43	4	5	6	8	9	10	11	13	16	18
44	4	5	6	7	9	10	11	13	15	18
45	4	5	6	7	8	9	11	13	15	17

A N N E X E C

.....

FICHES DE CALCUL ET DE RECUEIL DES DONNEES

DECRIRES DANS LE TEXTE

-oO()Oo-

IMPRIME 3.)

FICHE DE MESURE D'UNE PLACETTE-ECHANTILLON - PLANTATIONS

Page :

Nombre de pages remplies :

Forêt	Série	Parcelle	Placette n°	Essence
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Surface de la placette	Pente (en grades)	Date de la mesure:	Mois	Année
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Nombre de cartes (ne pas remplir)

	Arbre n°	Diamètre à 1,30 m sur écorce	Hauteur	Codes	
1ère carte	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2ème carte	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3ème carte	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4ème carte	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5ème carte	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

NOTES

.....

RESPONSABLE DE LA MESURE : DATE :/..../.....

FAIRE LES CALCULS DE HAUTEUR AU VERSO.

IMPRIME 5.1 : MODELES DE REGRESSION SIMPLE A PENTE UNIQUE OU A CONSTANTE UNIQUE

Fiche 1 : Récapitulation des données des placettes
(utiliser plusieurs fiches 1 si nécessaire)

Transformations des données utilisées :	X = _____
	Y = _____

Placette :

Données brutes		Données transformées		x ²	XY
		X	Y		
Totaux	n =	ΣX =	ΣY =	ΣX ² =	ΣXY =

Placette :

Données brutes		Données transformées		x ²	XY
		X	Y		
Totaux	n =	ΣX =	ΣY =	ΣX ² =	ΣXY =

Placette :

Données brutes		Données transformées		x ²	XY
		X	Y		
Totaux	n =	ΣX =	ΣY =	ΣX ² =	ΣXY =

: TRANSFORMATION DES DONNEES D'AGE ET DE HAUTEUR TIREES DE PLACETTES PERMANENTES POUR L'AJUSTEMENT DU PARAMETRE k DE L'EQUATION DE SCHUMACHER PAR UN ESTIMATEUR DE PENTE UNIQUE

i	j	A_{ij}	a	b	c	X_{ij}	H_{ij}	d	e	f	g	h	Y_{ij}
Placette	Mesure	Age	Somme	Diffé- rence	$\frac{2}{a}$	ln c	Hauteur	Somme	Diffé- rence	$\frac{e}{b}$	$\frac{d}{2}$	$\frac{f}{g}$	ln h

↘ A reporter sur la fiche 1 de l'imprimé 5.1

IMPRIME A1 : APPROXIMATION D'UNE COURBE PAR UN POLYNOME DU 2ème OU DU 3ème DEGRE -
CALCUL DES COEFFICIENTS

OBJECTIFS : Calculer les coefficients b_i de l'une ou l'autre des équations :

$y = b_0 + b_1x + b_2x^2 + b_3x^3$ (3ème degré : cubique)

$y = b_0 + b_1x + b_2x^2$ (2ème degré : parabole)

REMARQUE : Il est inutile de calculer les grandeurs marquées du signe \diamond si on se contente d'un polynôme du 2ème degré.

- ① Dresser un tableau contenant les coordonnées, ainsi que les carrés et les cubes des abscisses, de 3 ou 4 points de la courbe :

i	y_i	x_i	x_i^2	x_i^3
1				\diamond
2				\diamond
3				\diamond
4	\diamond	\diamond	\diamond	\diamond

- ② A partir du tableau ①, calculer les différences ci-dessous :

i	$y_{i+1} - y_i$	$x_{i+1} - x_i$	$x_{i+1}^2 - x_i^2$	$x_{i+1}^3 - x_i^3$
1				\diamond
2				\diamond
3	\diamond	\diamond	\diamond	\diamond

- ③ A partir du tableau ②, calculer les termes du tableau ci-dessous :

i	u_i	v_i	w_i
1			\diamond
2			\diamond
3	\diamond	\diamond	\diamond

avec : $u_i = (y_{i+1} - y_i) / (x_{i+1} - x_i)$

$v_i = (x_{i+1}^2 - x_i^2) / (x_{i+1} - x_i)$

$w_i = (x_{i+1}^3 - x_i^3) / (x_{i+1} - x_i)$

- ④ A partir du tableau ③, calculer les termes du tableau ci-dessous :

i	$v_{i+1} - v_i$	p_i	q_i
1			\diamond
2	\diamond	\diamond	\diamond

avec : $p_i = (u_{i+1} - u_i) / (v_{i+1} - v_i)$

$q_i = (w_{i+1} - w_i) / (v_{i+1} - v_i)$

- ⑤ Calculer les coefficients à l'aide des formules suivantes :

<u>Coefficient</u>	<u>Cubique (3ème degré)</u>	<u>Parabole (2ème degré)</u>
b_3	$(p_2 - p_1) / (q_2 - q_1)$	0
b_2	$p_1 - b_3 q_1$	p_1
b_1	$u_1 - b_2 v_1 - b_3 w_1$	$u_1 - b_2 v_1$
b_0	$y_1 - b_1 x_1 - b_2 x_1^2 - b_3 x_1^3$	$y_1 - b_1 x_1 - b_2 x_1^2$

IMPRIME A2 : CALCUL DE DIVERSES STATISTIQUES DANS LA REGRESSION LINEAIRE SIMPLE, EN UTILISANT UN IMPRIME DE DAWKINS (1968) REPRODUIT ICI AVEC LA PERMISSION DE L'AUTEUR.

Ligne	REGRESSION LINEAIRE et CORRELATION				
1	Nature des observations :				
2	Variable expliquée y :				
3					
4	Variable explicative x :				
5					
6	Sommes de carrés (SC) et de produits (SP)		n =	observations	
7	$(\Sigma y)^2 =$	$\Sigma y \times \Sigma x =$	$(\Sigma x)^2 =$	y	x
8	$\Sigma y^2 =$	$\Sigma yx =$	$\Sigma x^2 =$		
9	$-(\Sigma y)^2/n =$	$-\Sigma y \times \Sigma x/n =$	$-(\Sigma x)^2/n =$		
10	$SC_y =$	$SP_{yx} =$	$SC_x =$		
11	Régression		coefficient: $SP_{yx}/SC_x =$	= b	
12	Linéaire		constante: $\bar{y} - b\bar{x} =$	= a	
13	SC expliquée par la régression = $b \times SP_{yx} =$				
14	SC totale = SC_y ; nombre total de d.d.l. = $n-1 =$				
15	SC résiduelle et nombre résiduel de d.d.l. = obtenus par différence				
16	ANALYSE DE VARIANCE				
17	Source de variation	SC	d.d.l.	CM	F de la
18	Régression (expliquée)				observé
19	Résiduelle				table
20	Totale				(au seuil p=)
21	Carré du coefficient de corrélation, $r^2,$				
22	SC expliquée par la régression/SC totale = / = $-r^2$				
23	Coefficient de corrélation linéaire, r,				
24	$r = \sqrt{r^2} =$ avec le signe de b				
25	Ecart-type résiduel = $\sqrt{CM \text{ résiduel}} =$ = ET résiduel				
26	Coefficient de variation résiduelle en % = $100 \times ET \text{ résiduel} / \bar{y} =$ = CV% Σy Σx				
28	Le t de Student pour n-2 d.d.l. et P=0,05 est :				
				\bar{y}	\bar{x}
30	L'écart-type du coefficient b est $\sqrt{CM \text{ résiduel} / SC_x} =$ = ETb				
31	Son t-observé est $b/ETb,$ ou $\sqrt{F\text{-observé}} =$ = t-observé				
32	Intervalle de confiance de la moyenne y' de m valeurs de y estimées par la régression à partir de m valeurs de x (de moyenne x') :				
34	$y' \pm t \times ET \text{ résiduel} \times \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x' - \bar{x})^2}{SC_x}}$				
35					

IMPRIME A3 : CALCUL DES COEFFICIENTS ET DE DIVERSES STATISTIQUES D'UNE REGRESSION LINEAIRE A DEUX VARIABLES EXPLICATIVES

Partie 2

(10) CORRELATIONS ENTRE VARIABLES
$r_{yx} = \Sigma xy / \sqrt{(\Sigma x^2 \times \Sigma y^2)} =$
$r_{yz} = \Sigma yz / \sqrt{(\Sigma z^2 \times \Sigma y^2)} =$
$r_{xz} = \Sigma xz / \sqrt{(\Sigma x^2 \times \Sigma z^2)} =$

(11) COEFFICIENT DE CORRELATION MULTIPLE R, & R ²
$R^2 = \Sigma \hat{y}^2 / \Sigma y^2 =$
$R = \sqrt{R^2} =$

(12) ANALYSE DE VARIANCE	Somme des carrés	d.d.l.	Carré moyen	F-observé
Régression	$\Sigma \hat{y}^2 =$	2		
Résiduelle	$\Sigma d^2 =$	n-3 =		$= s^2$
Totale	$\Sigma y^2 =$	n-1 =		

(13) ECART-TYPE RESIDUEL
$s = \sqrt{s^2} =$

(14) ECART-TYPE ET t-OBSERVE DES COEFFICIENTS	b ₁	b ₂
Coefficient		
Ecart-type s _b	$s \times \sqrt{g_1} =$	$s \times \sqrt{g_3} =$
t = b/s _b		

(15) INTERVALLES DE CONFIANCE DES ESTIMATIONS
L'écart-type de la moyenne Y' de m valeurs de Y, estimées par la régression à partir de m valeurs de X et Z (de moyennes X' et Z'), est :
$s_{y'} = s \times \sqrt{(1/m + 1/n + g_1 x^2 + 2g_2 xz + g_3 z^2)}$
avec $x = (X' - \bar{X})$ et $z = (Z' - \bar{Z})$. Pour l'estimation d'un ensemble important d'individus, on peut faire $1/m = 0$; pour l'estimation d'un seul individu, $1/m = 1$.

A N N E X E D

BIBLIOGRAPHIE COMMENTEE

Cette bibliographie ne fournit pas une liste complète des ouvrages de dendrométrie. Elle propose seulement quelques ouvrages de référence auxquels on pourra se reporter pour approfondir ses connaissances sur les techniques mentionnées dans ce manuel.

Les références des articles techniques cités dans le texte ont fait l'objet de notes en bas de pages et ne sont pas reprises ici.

ASSMANN E. 1970

The principles of forest yield study - Pergamon press - 506 pages.

Traité très complet sur les aspects théoriques de la croissance forestière, illustré de nombreux exemples. Les principes généraux qui y sont exposés sont tirés de l'observation de peuplements équiennes d'Europe du Nord et ne s'appliquent pas nécessairement aux forêts tropicales. Seuls les modèles de production statiques et allométriques y sont abordés.

AVERY T.E. 1967

Forest measurements - Mac-Graw-Hill - 290 pages.

L'accent est mis sur les aspects pratiques du travail sur le terrain et de l'analyse préalable des données, plutôt que sur la construction des modèles. Se référant à des pratiques et à une terminologie nord-américaines ce livre n'est pas parfaitement adapté aux tropiques, mais n'en constitue pas moins un manuel utile.

BURLEY J. & WOOD P.J. [présenté par] 1976

A manual on species and provenance research with particular reference to the tropics. Tropical forestry paper 10, Commonwealth forestry Institute, Oxford.

Complémentaire du présent manuel, il traite des exigences pratiques de la recherche dans les études préalables à l'introduction d'essences exotiques en plantation. Concret et écrit dans un langage simple, il traite en annexe des études sur la qualité du bois (par J.F. Hughes et R.A. Plumtre).

CAB 1977

Computerized methods in forest planning and forecasting. Annotated bibliography F14 - Commonwealth agricultural bureau, United Kingdom.

Bibliographie de la littérature publiée dans le monde de 1973 à 1976, avec un résumé détaillé sous chaque référence. La plupart des sujets abordés relèvent de la prévision et du contrôle de la production.

CARRON L.T. 1968

An outline of forest mensuration - Australian national university press - 224 pages

Plus tourné que les textes américains ou européens vers la construction de tables de production et de tarifs de cubage dans les plantations à large espacement. Les méthodes y sont brièvement et clairement résumées mais les problèmes pratiques posés par la construction et l'ajustement des modèles n'y sont pas abordés.

DANIEL C. & WOOD F.S. 1971

Fitting equations to data - Wiley interscience - 342 pages

Traite de l'application pratique de la régression multiple à l'ajustement des modèles, y compris l'analyse des résidus, la régression sériée, la régression progressive ("stepwise regression") et l'ajustement de modèles non linéaires. Donne des exemples et de la documentation sur un ensemble de programmes, mais ne donne le texte d'aucun programme.

DAWKINS H.C. 1968

STATFORMS - Formats for elementary statistical calculation - Commonwealth Forestry Institute, Oxford - Institute paper n° 41

26 feuilles pour les calculs manuels nécessaires dans des techniques statistiques telles que l'analyse de variance (carré latin) et de covariance, la régression linéaire, l'échantillonnage aléatoire stratifié, ...etc, indiquant les différentes étapes du calcul. Très utile pour de petits calculs manuels ; exemples utiles.

DAY A.C. 1972

FORTRAN techniques - Cambridge University Press - 96 pages

Bon supplément à un cours d'initiation au FORTRAN. Présente des méthodes de tracé de graphiques, de tri, d'indexation, d'utilisation de sous-programmes adressés, ouverts ou récursifs, ...etc.

DRAPER N. & SMITH H. 1966

Applied regression analysis - Wiley-Interscience - 407 pages

Fait presque figure de "classique" de la régression multiple et de la régression non linéaire, avec au sommaire : la théorie de la régression simple, la formulation matricielle de la régression multiple, l'analyse des résidus, la régression progressive, l'utilisation de la régression pour l'analyse des résultats d'expériences contrôlées et une introduction à la régression non linéaire. Sa lecture demande des connaissances en algèbre matricielle et en calcul différentiel.

FAO 1975

Manuel d'inventaire forestier avec références particulières aux forêts tropicales hétérogènes - Food and Agriculture Organization - Rome - 200 pages [en français]

Référence essentielle pour les études de production. Traite longuement des dispositifs d'échantillonnage en forêt et des formules pour l'exploitation des résultats. Guide pratique pour la réalisation d'un inventaire forestier et l'enregistrement des données.

FRIES J. [présenté par] 1974

Growth models for tree and stand simulation - Research note n°30, Department of forest yield research, Royal College of Forestry, Stockholm - 379 pages - (Disponible auprès de J. FRIES, University of agricultural Science, S-75007 UPPSALA, Suède, pour 25 \$US).

Recueil de communications aux réunions du sous-groupe de travail S4.01.4 de l'IUFRO. Beaucoup d'idées utiles.

FRIES J., BRUKHART H., MAX T. [présenté par] 1978

Growth models for long term forecasting of timber yield - FWS.1.78, School of forestry and wildlife resources, Virginia polytechnic institute and State university, BLACKSBURG, VA 24061, USA - 249 pages.

Recueil des communications présentées à une réunion du groupe de travail S4-01 de l'IUFRO. Beaucoup d'idées utiles.

GREEN P.E. & CARROLL J.D. 1976

Mathematical tools for applied multivariate analysis. Academic press. - 376 pages.

Un bon livre pour apprendre seul l'algèbre matricielle appliquée à la régression multiple, à l'analyse en composantes principales, à l'analyse factorielle des correspondances et à la classification.

HUSCH B., MILLER C.I. & BEERS T.W. 1972

Forest mensuration - Ronald press, New-York - 410 pages.

Un "classique" de la dendrométrie, qui met l'accent sur les mesures plutôt que sur les modèles et n'aborde que très peu la prévision de la croissance. Il se réfère aux conditions nord-américaines = certaines des pratiques présentées ne sont donc pas transposables aux forêts tropicales. Manuel utile pour acquérir les bases de la dendrométrie.

LAND A. & POWELL S. 1973

FORTTRAN codes for mathematical programming - Wiley - 249 pages

Présente des programmes fonctionnels, avec une documentation complète, pour la programmation mathématique : linéaire, en nombres entiers, ou quadratique ; explication rapide de la théorie sous-jacente.

PARDE J. 1961

Dendrométrie - Ecole Nationale des Eaux et Forêts, Nancy - 350 pages [en français]

Encore un classique de la dendrométrie, qui met un peu plus l'accent que HUSCH et alters sur la construction des tables de production. Très bonne présentation de l'utilisation des instruments de mesure (du relascope de BITTERLICH entre autres) ; des indications intéressantes sur la conception de modèles simples traduisant les inter-relations entre variables d'un peuplement. C'est sans doute le meilleur manuel de dendrométrie existant actuellement, mais il n'aborde pas les problèmes tropicaux.

PIELOU E.C. 1977

Mathematical ecology - Wiley Interscience - 385 pages.

Excellent ouvrage de référence sur la modélisation de la dynamique des populations dans le temps et dans l'espace, et sur l'analyse des associations végétales. Plus intéressant pour le sylviculteur-écologiste que pour le gestionnaire ; le chapitre qui traite de l'analyse des associations végétales est important pour les études en forêt tropicale humide.

POOLE R.W. 1974

An introduction to quantitative ecology - Mac-Graw-Hill - 532 pages.

Présente un large éventail des techniques utilisées en écologie, en mettant l'accent sur la construction de modèles. Il peut constituer une bonne introduction au sujet car, moins théorique que l'ouvrage de PIELOU, il offre une information plus élémentaire avec des exemples plus nombreux ; il suppose quand même quelques connaissances en algèbre (matrices) et analyse (calcul différentiel).

PRODAN M. 1968

Forest biometrics - Pergamon Press - 447 pages.

Exposé très utile de l'arrière-plan statistique des études de croissance en forêt, qui n'exige qu'un niveau moyen de connaissances mathématiques. Il expose bien l'ajustement des fonctions non linéaires, mais sans souci direct d'application pratique. Se réfère essentiellement à des pratiques forestières européennes.

ROYCE-SADLER D. 1975

Numerical methods for non-linear regression - University of Queensland Press - 89 pages.

C'est un résumé des principales méthodes utilisées dans l'ajustement des modèles non-linéaires, qui suppose certaines connaissances en algèbre et analyse. Exposé très clair et concis des algorithmes essentiels.

SEBER G.A.F. 1977

Linear regression analysis - Wiley Interscience - 465 pages.

Plus théorique, plus complet et plus à jour que l'ouvrage de DRAPER & SMITH. Livre de référence très utile au statisticien, il requiert une bonne compréhension de l'algèbre matricielle.

SHANNON R.E. 1975

Systems simulation : the art and science - Prentice Hall - 387 pages.

Ouvrage de référence utile pour ceux qui s'intéressent à la construction de modèles de simulation : génération d'un nombre au hasard, ajustement des modèles, analyse des systèmes, expériences sur modèle, décision sur modèle ; contient plusieurs études de cas.

SNEDECOR G.W. & COCHRAN W.G. 1967

Statistical methods - Iowa state university press - 593 pages.

Excellent ouvrage de référence sur les méthodes statistiques, il ne demande pas de trop grandes connaissances mathématiques (on peut ignorer l'algèbre matricielle et l'analyse). Il traite, entre autres sujets, de la régression linéaire, non-linéaire, multiple, des dispositifs expérimentaux et d'échantillonnage, et de leur analyse. De nombreux exemples illustrent les différentes techniques et chaque chapitre est suivi d'une série de petits exercices permettant d'en vérifier la compréhension.

STARK P.A. 1970

Introduction to numerical methods - Mac Millan - 334 pages.

Ouvrage fondamental sur les méthodes de calcul numérique pour ordinateur : résolution des équations et des systèmes d'équations, intégration et différenciation numériques, interpolation polynomiale. La plupart des méthodes présentées sont illustrées de courts programmes en FORTRAN. Le niveau mathématique requis est celui de l'entrée à l'Université : le livre donne des notions d'algèbre matricielle mais suppose acquis un certain niveau en analyse.

UNIVERSAL ENCYCLOPAEDIA OF MATHEMATICS - Pan books - 715 pages.

Ouvrage de référence peu coûteux (édition de poche) et très utile, donnant des formules d'algèbre, trigonométrie, analyse, etc... , ainsi que des tables mathématiques et un lexique.

WAGNER H.M. 1975

Principles of management science - Prentice Hall - 612 pages.

Excellent ouvrage sur les méthodes quantitatives de décision dans la gestion : méthodes de programmation linéaire et mathématique, gestion des stocks, modèles de simulation et de files d'attente. Ne requiert pas de connaissances mathématiques particulières.

WRIGHT H. [présenté par] 1980

Planning performance and evaluation of growth and yield studies - Commonwealth Forestry Institute, Oxford, Great-Britain - [Sous-presse]

Recueil des communications présentées à une réunion du groupe de travail S4.01 de l'IUFRO - Beaucoup d'idées utiles.