

Dépérissement des forêts : essai d'analyse des dépendances

D. MANDALLAZ, R. SCHLAEPFER *, J. ARNOULD **

Chaire d'aménagement des forêts, EPF Zürich

** Chaire d'aménagement des forêts*

Institut pour la Recherche sur la Forêt et le Bois de l'Ecole Polytechnique Fédérale de Zürich

*** GREF*

Résumé

Le présent article expose une approche méthodologique pour l'analyse des interdépendances entre l'état de santé de l'épicéa, du sapin et du hêtre et certaines caractéristiques tant stationnelles que propres à l'arbre individuel. Il s'appuie sur plusieurs inventaires par placette effectués en Suisse et tout particulièrement sur celui réalisé dans les forêts de la commune de Sainte-Croix (canton de Vaud, Suisse).

L'analyse statistique repose sur l'utilisation d'un modèle de régression logistique ; celui-ci met en relation la probabilité pour un arbre d'être déclaré sain (la distinction entre « sain » et « non sain » se faisant au seuil de 20 p. 100 de perte — ou, plus exactement, de manque — de feuillage) avec plusieurs facteurs comme le pH et la région — en ce qui concerne les variables stationnelles — ou le diamètre, l'essence, la position sociale et la longueur du houppier — en ce qui concerne les arbres individuels —. L'élaboration du modèle finalement retenu s'est déroulée en deux phases, exploratoire puis confirmatoire, cette dernière ayant reçue une attention toute particulière. La probabilité d'être déclaré sain apparaît avant tout comme une fonction monotone décroissante du diamètre ; mais elle dépend également, de façon plus complexe, de plusieurs autres facteurs, dont l'essence, la position sociale et la longueur du houppier (une bonne position sociale et une longue cime influencent en général favorablement l'état sanitaire).

Un tel modèle est susceptible d'être complété par d'autres paramètres décrivant certaines caractéristiques chimiques foliaires ou atmosphériques.

1. Introduction

La description objective de l'état de santé de la forêt, à un moment donné et dans son évolution, de même que l'analyse des interdépendances dans l'écosystème, font partie des tâches les plus importantes des forestiers dans le contexte du dépérissement des forêts (SCHLAEPFER *et al.*, 1985).

R. SCHLAEPFER, D. MANDALLAZ, Institut pour la Recherche sur la Forêt et le Bois, ETH Zentrum, CH 8092 Zürich.

Dans cette optique, la chaire d'aménagement des forêts de l'EPFZ a effectué plusieurs inventaires dans différentes entreprises ⁽¹⁾ forestières et supervisé l'analyse statistique d'études à grande échelle, ceci dans le cadre du programme suisse « Sanasilva ».

Le but de cet article est de présenter les résultats préliminaires de nos recherches. Nous mettrons particulièrement en évidence les problèmes méthodologiques de validation des modèles, en prenant comme exemple un inventaire effectué dans les forêts communales de Sainte-Croix (Canton de Vaud).

En certaines occasions, nous placerons le problème dans un contexte plus général, en rapportant nos expériences sur un nombre de cas certes trop limité pour avancer des conclusions définitives, mais suffisant pour apprécier la complexité du problème.

2. Le domaine d'enquête

Le domaine d'étude, situé dans la zone du Jura plissé, repose sur des couches géologiques superposées du Jurassique et du Crétacé, auxquelles viennent s'ajouter, en zone basse, des structures morainiques ; le sol est essentiellement composé de sols bruns, de sols bruns lessivés et de rendzines là où le sous-sol calcaire affleure.

L'altitude des placettes varie entre 1 040 et 1 280 m, les peuplements sont donc soumis à un climat montagnard, avec des précipitations d'environ 1 500 mm (la partie Nord-Est du domaine a une nette orientation Nord/Nord-Est), une période de végétation de 120 à 170 jours et des températures moyennes annuelles de 2 °C à 7 °C.

Quelques remarques doivent être faites aux niveaux forestier et sylvicole :

En premier lieu, à de telles altitudes et à une telle exposition nord, se trouvent les zones limites de la hêtraie à sapins et commence la zone d'extension de l'épicéa. En second lieu, la gestion forestière rencontre ici des difficultés, tant par les séquelles d'un problème sylvo-pastoral antérieur, que par le net déséquilibre des classes d'âge au détriment des plus jeunes. Cette situation se manifeste dans l'instabilité des peuplements face aux agressions climatiques et aux interventions sylvicoles.

La surface forestière soumise au régime forestier est de 1 715 ha, dont 364 ha pour la série inventoriée dans le cadre de l'étude. 87 % du volume sur pied est constitué de conifères (67 % d'épicéas, 20 % de sapins), les 13 % restants correspondent essentiellement à du hêtre.

3. L'inventaire

3.1. Méthode

La méthode retenue consiste en un inventaire systématique par placette, suivant un réseau à mailles de 200 × 300 m (soit une placette pour 6 ha). Seules sont

(1) Le terme d'entreprise semble spécifique à la Suisse Romande et correspondrait, en France, à une « exploitation forestière » au sens où l'on parle d'« exploitation agricole » ; il s'agit donc d'une ou de plusieurs unités de gestion.

retenues les placettes dont les arbres ont un diamètre supérieur à 12 cm. Les 15 arbres les plus proches du centre de la placette sont pris dans l'échantillon.

L'inventaire comprend 45 placettes, dont 11 sont reprises de l'inventaire vaudois de 1969, ce qui a permis une étude de l'influence de la perte de feuillage sur l'accroissement (SCHLAEPFER, 1985). Les caractéristiques suivantes sont relevées :

Pour la placette :

Type et structure du peuplement, stade d'évolution, degré de fermeture (ouvert ou fermé) et de mélange, altitude, exposition, pente, relief, 2 mesures de pH à 10 cm de profondeur, 2 mesures de profondeur du sol.

Pour l'arbre :

Essence, diamètre à hauteur de poitrine (dhp), position sociale, longueur du houppier (en p. 100 de la hauteur totale), p. 100 de perte — ou, plus exactement, de manque — de feuilles ou d'aiguilles et autres caractéristiques de l'état de santé (dégâts biotiques et abiotiques, degré de jaunissement : ces grandeurs n'ont présenté aucun intérêt statistique et ne seront pas étudiées ici).

3.2. Présentation générale de l'état de santé

Les observations portent sur 675 arbres, dont 101 hêtres (15 p. 100), 254 sapins (37 p. 100) et 314 épicéas (47 p. 100), en futaie le plus souvent fermée. L'âge des peuplements a été évalué, pour 80 p. 100 d'entre eux, à plus de 110 ans.

Les figures 1, 2, 3 donnent pour les 3 essences la distribution des pertes de feuillage.

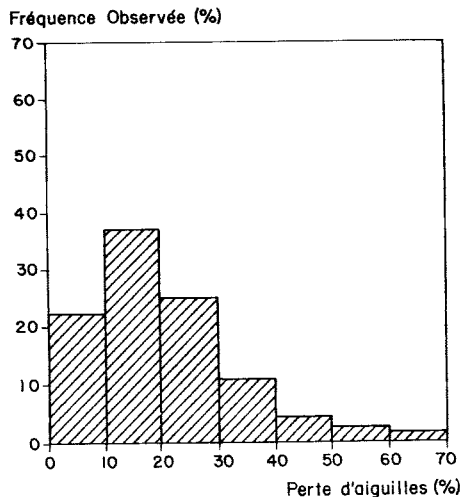


FIG. 1

*Distribution de la perte d'aiguilles chez l'épicéa.
Distribution of foliage loss in spruce.*

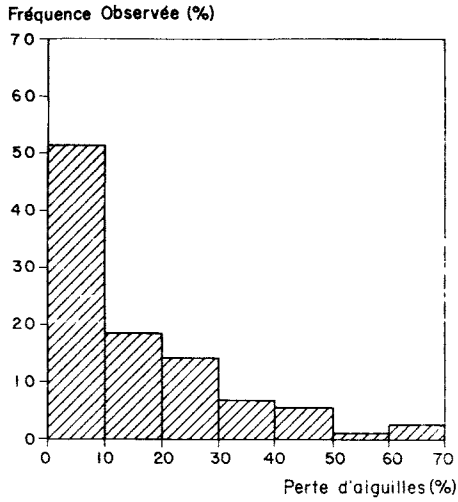


FIG. 2

Distribution de la perte d'aiguilles chez le sapin.
Distribution of foliage loss in fir.

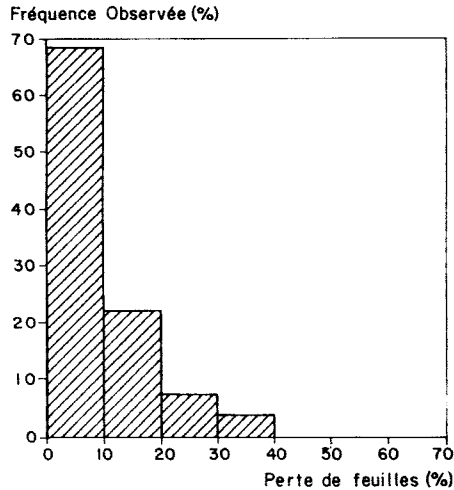


FIG. 3

Distribution de la perte de feuilles chez le hêtre.
Distribution of foliage loss in beech.

Pour des raisons traitées en détail au chapitre IV nous donnons déjà la répartition des arbres selon leur classification « sain » et « non-sain », basée sur un seuil de perte de feuillage fixé à 10 p. 100 ou 20 p. 100 (tabl. 1).

TABLEAU I

Etat de santé général.
Overall state of health.

Seuil		Hêtre	Epicéa	Sapin
10 %	Proportion d'arbres sains	67 %	23 %	50 %
	Erreur standard	6,8 %	5,3 %	6,4 %
20 %	Proportion d'arbres sains	90 %	60 %	70 %
	Erreur standard	3,6 %	5,2 %	6,7 %

Les erreurs standards sont calculées selon la méthode de COCHRAN (1977) pour l'échantillonnage en grappes.

4. Analyse par un modèle de régression logistique

4.1. Préliminaires théoriques

Y_{ij} est une variable aléatoire binaire prenant la valeur 0 si l'arbre i de la placette j est sain, 1 s'il ne l'est pas.

a_j est un vecteur de variables explicatives qui décrivent la placette (relief, exposition, pH, etc.).

b_{ij} est un vecteur de variables explicatives qui décrivent l'arbre particulier (essence, diamètre, position sociale, longueur du houppier, etc.).

La probabilité que l'arbre soit sain est supposée être de la forme :

$$p_{ij} = \Pr(Y_{ij} = 0/a_j, b_{ij}) = \exp(a_j \times \beta + b_{ij} \times \tau) / [1 + \exp(a_j \times \beta + b_{ij} \times \tau)]$$

où β , τ dénotent des vecteurs de paramètres inconnus et \times le produit scalaire.

Le lecteur pourra consulter l'excellent ouvrage de D.R. Cox (1980) pour une étude détaillée de l'analyse des données binaires par le modèle logistique, méthode désormais classique.

Notons que l'analyse est basée entièrement sur des probabilités conditionnelles. Les Y_{ij} sont supposés être indépendants conditionnellement aux covariables observées, ce qui permet de prendre en compte partiellement certains effets de compétition entre arbres.

La transformation logistique

$$l_{ij} = \log [p_{ij}/(1 - p_{ij})] = a_j \times \beta + b_{ij} \times \tau$$

et l'utilisation du maximum de vraisemblance pour l'estimation de β et τ permettent de ramener formellement le problème aux techniques usuelles de régression multiple : chaque composante des vecteurs β , τ décrit l'influence du facteur associé en gardant

les autres constants. Notons que, contrairement aux problèmes habituels d'estimation en échantillonnage, les probabilités d'inclusion dans l'échantillon n'apparaissent pas explicitement dans le modèle (dans le cas présent celles-ci sont d'ailleurs inconnues).

Pour l'interprétation, il suffit de noter que p_{ij} est une fonction croissante de I_{ij} : les autres facteurs restant constants, si le paramètre estimé d'un effet est positif, la probabilité que l'arbre soit « sain » augmente dès lors que cet effet croît (pour les variables indicatrices 0/1, « croît » signifie : « passe de 0 à 1 »), et inversement si le paramètre est négatif. Cette approche peut avoir un caractère fictif dans la mesure où ceci n'est possible que pour des petites variations des paramètres : la distinction entre variables « explicatives » et variables « dépendantes » relève en effet plus de la statistique que de la réalité. Remarquons finalement que les données analysées sont purement « observationnelles » et ne correspondent pas du tout aux plans d'expérience « optimaux » : la structure factorielle est déséquilibrée et ne permet pas l'estimation d'interactions complexes. Nous émettons donc les réserves d'usage, en précisant bien qu'elles s'appliquent également à toutes les autres méthodes d'analyse.

4.2. Définition d'un arbre sain

L'analyse repose sur la dichotomisation des états de santé entre arbre « sain » et arbre « non sain » (ou « malade »). Cette simplification draconienne nous paraît justifiée pour les raisons suivantes :

- a) Le problème est d'une telle complexité qu'une approche par étapes s'impose. La taille des échantillons étant généralement très grande, les effets potentiels doivent apparaître même sur des classifications simplifiées.
- b) L'état de santé est une grandeur ordinaire et partiellement subjective (la formation commune des observateurs ne garantissant que l'homogénéité). Des analyses « classiques » par régression multiple sur les p. 100 de perte de feuillage ne sont ni légitimes, au sens strict, ni plus simples, dès lors qu'on est confronté à des tests de validation sur des résidus non gaussiens.
- c) Le nombre d'arbres dans les catégories supérieures de perte de feuillage est (encore ?) faible.
- d) Le modèle logistique peut être facilement adapté à 3 ou 4 classes d'état de santé, par probabilités conditionnelles successives dans les « non sains ».
- e) La dichotomisation peut se baser sur une réduction de données multivariées décrivant l'état de santé.

La simplification n'intervient qu'au stade de l'analyse : l'observation elle-même doit être aussi fine que la précision et les autres contraintes de l'inventaire le permettent.

Généralement, la classification « sain/non sain » repose sur le p. 100 de perte de feuilles ou d'aiguilles, le seuil classiquement retenu étant de 10 p. 100. Il nous semble qu'un seuil de 20 p. 100 est plus robuste et conduit, généralement, à de meilleurs ajustements. De plus il n'est pas biologiquement établi qu'une différence de 10 p. 100 par rapport à un arbre dont la cime est « indiscutablement bien fournie » puisse être considérée comme l'expression d'un état maladif, voire malade.

Dans le cas de Sainte-Croix, les 2 seuils ont été utilisés, avec des résultats équivalents, pour l'essentiel, quant à l'interprétation. Dans cet article, nous ne présentons que l'analyse avec 20 p. 100, et un arbre sera donc considéré comme « malade » si sa perte de feuilles ou d'aiguilles est estimée à plus de 20 p. 100.

4.3. Présentation du modèle logistique retenu

Méthodologiquement, il importe de bien distinguer deux phases de l'analyse :

Dans une première phase, *exploratoire*, un sous-échantillon aléatoire simple de 336 arbres, tirés parmi les 652 arbres inventoriés retenus pour l'analyse (arbres avec observations complètes parmi les 669 hêtres, épicéas et sapins), a servi à l'élaboration du modèle (échantillon de construction). En se basant d'une part sur des études antérieures, d'autre part sur plusieurs modèles préliminaires ainsi que sur des critères de sélection biologiques, statistiques, numériques et esthétiques (simplicité), nous avons retenu un modèle incorporant les facteurs suivants : région, pH superficiel, profondeur du sol, degré de fermeture, essence, diamètre à hauteur de poitrine (dhp), position sociale et longueur du houppier chez le sapin seulement.

La variable « région » prend globalement en compte et, de façon purement qualitative, l'altitude, l'exposition et le relief, qui diffèrent significativement pour les trois régions retenues, contrairement au pH superficiel et à la profondeur du sol. Ces trois régions correspondent par ailleurs à des domaines géographiques bien distincts.

Dans une deuxième phase, *confirmatoire*, le modèle retenu est testé sur l'échantillon de confirmation (complémentaire de l'échantillon de construction dans l'ensemble des 652 arbres, soit 316 arbres). Cette approche évite les difficultés inhérentes à une vérification des modèles sur des données ayant servi à leur élaboration.

TABLEAU 2

*Résumé de l'analyse par le modèle logistique
Valeurs pour l'échantillon complet de 652 arbres
(valeurs pour l'échantillon de construction).*

*Summary of analysis by logistic regression
values for complete sample of 652 trees (values for exploratory sample).*

Effets	Degrés de liberté	Chi-deux	Seuil observé
Constante	1	37,4 (18,4)	0,0001
Région	2	25,7 (10,8)	0,0001
Profondeur	1	4,4 (2,2)	0,035
pH superficiel	1	6,4 (5,2)	0,012
Degré de fermeture	1	15,6 (18,1)	0,0001
Essence	2	2,5 (1,5)	0,29
Diamètre par essence	3	63,2 (31,5)	0,0001
Position sociale chez le sapin	4	25,0 (11,4)	0,0001
Longueur du houppier chez le sapin ..	1	10,6 (3,8)	0,001
Profondeur * essence	2	1,3 (3,2)	0,52

Toutefois, les graphiques et tableaux présentés sont établis à partir de l'échantillon complet. Les tableaux 2 et 3 donnent les résultats obtenus et contiennent à titre comparatif les valeurs obtenues sur l'échantillon de construction.

Pour certains paramètres (pH, position sociale, profondeur), la puissance des tests n'est satisfaisante que pour des échantillons relativement grands (500 arbres environ selon les simulations).

Pour de très grands échantillons (le plus grand à ce jour comprend 12 298 arbres), une séparation nette entre phase exploratoire et phase confirmatoire ne semble pas être absolument indispensable, bien que toujours recommandable.

Remarques :

- 1) Une analyse stratifiée par essence conduit qualitativement aux mêmes résultats.
 - 2) Notons l'absence d'influence de la position sociale et du houppier sur l'état de santé du hêtre et de l'épicéa. Ce résultat ne peut être généralisé.
 - 3) L'absence d'influence du diamètre chez le hêtre pourrait être dû au faible effectif (cf. tableau 3 ci-dessous). En effet, d'autres études révèlent également un effet négatif significatif du diamètre chez le hêtre, comme chez le sapin et l'épicéa.
 - 4) Le modèle présenté est en quelque sorte minimal : l'interaction « profondeur * essence » n'est retenue qu'à titre illustratif ; l'effet principal « essence » — non significatif — est nécessaire de par la structure semi-hiérarchique du modèle.
 - 5) La stabilité numérique des estimateurs sur des sous-échantillons est excellente (cf. tableau 3 ci-dessous). La convergence de l'algorithme de Newton-Raphson pour le maximum de vraisemblance est atteinte en 5-6 itérations. Les calculs ont été effectués avec le logiciel SAS (1982) sur une IBM 3083 de l'université de Zürich.
 - 6) A titre indicatif, le test du maximum de vraisemblance pour le modèle est de 558 avec 622 degrés de liberté (seules 640 configurations différentes des covariables ont été observées). Malheureusement la distribution asymptotique en chi-deux n'est pas légitime dans ce cas (le modèle serait excellent). Les tests chi-deux pour les effets isolés peuvent s'interpréter comme différences de déviations et sont donc asymptotiquement valides (McCULLAGH & NELDER, 1985 ; WILLIAMS, 1983).
- Avant de donner l'interprétation de ces résultats, nous examinons la validité du modèle.

5. Vérification du modèle

5.1. Généralités

Deux approches complémentaires sont possibles :

a) Au niveau de l'arbre individuel

Considérons les déviations d_{ij} définies par :

$$\begin{aligned} d_{ij} &= \left| -2 \times \log(1 - p^*_{ij}) \right|^{1/2} & \text{si } Y_{ij} = 1 \\ d_{ij} &= \left| -2 \times \log(p^*_{ij}) \right|^{1/2} & \text{si } Y_{ij} = 0 \end{aligned}$$

où p^*_{ij} dénote la probabilité estimée par le modèle logistique.

TABLEAU 3

Estimation des paramètres du modèle logistique (chiffres arrondis à la deuxième décimale).
Valeurs pour l'échantillon complet de 652 arbres (valeurs pour l'échantillon de construction).

Parameter estimates of logistic model (rounded to 2 decimals).
Values for complete sample of 652 trees (values for exploratory sample).

Paramètres	Estimateurs	Erreur standard	Seuil
Constante	6,47 (5,90)	1,06	0,0001
Région			
A	0,75 (0,72)	0,19	0,0001
B	0,04 (—0,11)	0,18	0,80
C	—0,79 (—0,61)	(1)	
Profondeur (2)	—0,03 (—0,03)	0,01	0,04
pH superficiel (2)	—0,38 (—0,43)	0,15	0,01
Fermeture			
fermé	0,53 (0,77)	0,13	0,001
ouvert	—0,53 (—0,77)	(1)	
Essence			
hêtre	—0,41 (—0,56)	0,66	0,90
épicéa	—0,78 (—0,70)	0,28	0,72
sapin	1,29 (1,25)	(1)	
Sapin			
prédominant	2,45 (1,92)	0,92	0,01
dominant	1,64 (1,76)	0,56	0,01
codominant	—0,04 (—0,48)	0,43	0,92
dominé	—1,30 (—0,89)	0,48	0,01
surcimé	—1,65 (—2,00)	(1)	
Diamètre (2)			
hêtre	0,01 (0,03)	0,04	0,98
épicéa	—0,05 (—0,05)	0,01	0,0001
sapin	—0,16 (—0,13)	0,03	0,0001
Houppier sapin (2)	0,04 (0,03)	0,01	0,01
Prof. * Ess.			
hêtre	—0,02 (—0,02)	0,02	0,50
épicéa	0,02 (0,03)	0,02	0,45
sapin	0,00 (—0,01)	(1)	

(1) Dans les effets à catégories discrètes (par exemple : « région ») le dernier paramètre est égal à — (somme des autres).

(2) Dans les effets continus (par exemple : « diamètre ») seul le terme linéaire est considéré (les termes quadratiques ne sont pas significatifs). Par exemple la grandeur de l'effet diamètre chez le sapin est de —0,16 par cm de dhp.

Des méthodes graphiques d'inspection des déviations d_{ij} en fonction des variables explicatives et des probabilités ajustées p^*_{ij} sont extrêmement utiles (McCULLAGH & NELDER, 1985). Les tests formels proposés dans la littérature (PREGIBON, 1981 ; WILLIAMS, 1983) ne sont malheureusement pas suffisamment sûrs quant à leur comportement asymptotique (ce point fera l'objet d'une publication ultérieure) et nous ne nous y attarderons pas.

b) *Au niveau de la placette*

Le modèle logistique déterminé par l'échantillon de construction permet de prédire le nombre attendu d'arbres malades dans l'échantillon de confirmation.

Soit en effet :

$$E_j = \sum (1 - p^*_{ij})$$

où i parcourt l'ensemble V_j des arbres de l'échantillon de vérification dans la placette j .

E_j est donc le nombre attendu d'arbres malades dans V_j , en supposant les covariables connues.

Soit O_j le nombre observé d'arbres malades dans V_j . Si le modèle est correct, O_j est une variable aléatoire de variance conditionnelle :

$$\text{Var}(O_j) = \sum p_{ij} \times (1 - p_{ij}) \quad (i \text{ dans } V_j)$$

et peut être estimée par :

$$\widehat{\text{Var}}(O_j) = \sum p^*_{ij} \times (1 - p^*_{ij}) \quad (i \text{ dans } V_j)$$

Il est donc naturel de considérer la statistique :

$$X^2_j = (O_j - E_j)^2 / \widehat{\text{Var}}(O_j)$$

En se référant au théorème central limite, on est en droit d'attendre, si le modèle est correct, que la statistique X^2_j suive une distribution de chi-deux avec 1 degré de liberté.

Donc :

$$X^2 = \sum X^2_j \quad j \text{ parcourant les placettes}$$

est, sous les mêmes hypothèses, distribué asymptotiquement comme un chi-deux avec N degrés de liberté, où N est le nombre de placettes.

Notons que, grâce à la procédure en deux phases, nous nous ramenons au cas de l'adéquation de fréquences observées à des fréquences données *a priori*, et le problème de la détermination des degrés de liberté ne se pose plus. De plus, le test conditionnel proposé est beaucoup plus puissant que le test chi-deux classique basé sur :

$$(O_j - E_j)^2 / E_j$$

Une analyse des résidus :

$$r_j = \text{signe}(O_j - E_j) \times (X^2_j)^{1/2},$$

en fonction de paramètres géographiques, est souvent très instructive.

5.2. Résultats pour Sainte-Croix

Le X^2 basé sur l'échantillon de confirmation est de 69 avec 44 degrés de liberté ($p < 0,01$). Toutefois, 2 placettes sont manifestement « aberrantes », et cela sans raisons apparentes. Le X^2 après leur exclusion tombe à 53 avec 42 degrés de liberté (n.s.), les paramètres restant pratiquement inchangés. Tenant compte du fait que, dans de grands échantillons, tout test d'ajustement devient significatif, nous pouvons admettre que le modèle reflète qualitativement assez bien la réalité. Ni une analyse graphique des déviations individuelles, ni une analyse des résidus dans les placettes n'ont permis de déceler une structure résiduelle évidente suggérant une modification du modèle. Il est intéressant de noter que le test de X^2 obtenu sur l'échantillon complet est de 73, et 54 après exclusion des 2 placettes « aberrantes ». Il semble donc, et ceci est confirmé par d'autres analyses, que les échantillons soient suffisamment grands pour compenser la perte de degrés de liberté due à l'estimation des paramètres (KENDALL & STUART, 1979). La distribution des r_j ne diffère pas significativement d'une distribution standard normale (test de Shapiro WILKS pour l'échantillon complet et l'échantillon de confirmation). A titre illustratif, la figure 4 donne l'histogramme des r_j pour l'échantillon complet. Notons que les tests proposés par PREGIBON (1981) et basés sur les déviations individuelles concluent à une parfaite concordance entre modèles et données, ce qui est trop optimiste dans le cas présent et parfaitement erroné dans d'autres analyses. En l'état actuel de nos connaissances, la méthode de prédiction au niveau des placettes semble être la seule à permettre l'emploi de tests d'ajustement dignes de confiance.

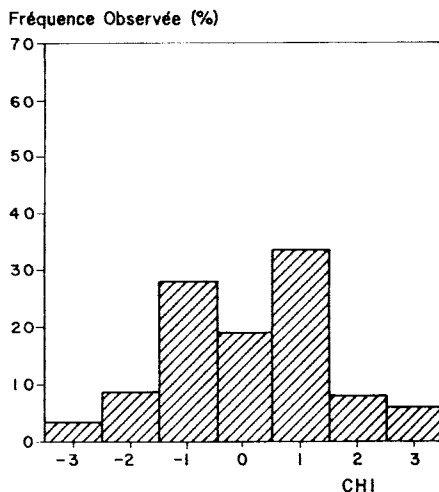


FIG. 4

Histogramme des résidus standardisés.
(Sainte-Croix, 44 placettes)

Histogram of standardized residuals.
(Sainte-Croix, 44 plots)

5.3. *Réflexions sur les autres études effectuées*

De façon générale, l'ajustement est qualitativement acceptable pour les inventaires au niveau de l'entreprise. Le nombre de placettes « aberrantes » varie entre 2 p. 100 et 10 p. 100, les paramètres restant très stables après leur exclusion (dans certains cas il est possible d'apporter une explication plausible au caractère « anormal » de ces placettes).

Relevons que les paramètres jugés significatifs ne sont pas toujours les mêmes, mais que nous n'avons pas trouvé de tendances significatives et de signes opposés selon les études.

Une autres étude, de grande envergure (Canton de Zürich : 170 000 ha dont 48 000 de forêts, 468 placettes, 12 298 arbres), approche de sa conclusion (MANDALLAZ, 1985 ; EICHENBERGER, 1985). Les premiers résultats fournis par un modèle incorporant 25 facteurs (ce qui nécessite 74 paramètres) indique, par une analyse géographique des résidus, que des variables exogènes décrivant le sol et l'atmosphère devraient être incorporées pour améliorer l'ajustement (X^2 de 1 402 avec 468 D.L. et de 710 après exclusion de 26 placettes « aberrantes »). L'interprétation est plus délicate dans la mesure où le diamètre, qui joue un rôle négatif pour toutes les essences (sauf le pin) au seuil de 10 p. 100, n'est significatif que pour l'épicéa et le hêtre au seuil de 20 p. 100. Des études complémentaires en cours, décrivant d'une part les pollutions atmosphériques (SO₂, NO_x), d'autre part les caractéristiques chimiques de la masse foliaire (teneur en S, N, Ca, K, etc.), permettront de compléter le modèle. De plus, le premier modèle reposait sur un découpage en partie administratif des paramètres « régionaux », ce qui doit être modifié.

Une explication plausible à ces difficultés serait que les variables exogènes, de par leur variabilité, jouent un rôle important, voire primordial, sur des grandes surfaces, alors qu'elles pourraient être négligées sur des petites, où elles peuvent être considérées constantes.

D'un point de vue pratique, il est utile de noter qu'une étude de cette ampleur requiert des moyens de calcul importants (40 mn de CPU par modèle sur une IBM 3083). Des investigations à plus grande échelle encore devraient employer la méthode des moindres carrés pondérés itératifs, ce qui implique une certaine simplification des modèles, notamment un regroupement du dhp en classes.

6. Interprétation des résultats

6.1. *Diamètre et essence*

A Sainte-Croix, le diamètre influence très fortement l'état de santé du sapin et de l'épicéa. La probabilité d'être malade est, chez le sapin encore plus fortement que chez l'épicéa, une fonction monotone croissante du diamètre (et donc, corrélativement, de l'âge); toutefois, la variation des autres facteurs estompe les différences entre sapin et épicéa au niveau global (cf. tabl. 1). En ce qui concerne le

hêtre, contrairement aux résultats obtenus à Zürich, une telle dépendance n'a pu être décelée : ceci tient probablement à la petite proportion d'arbres malades dans l'effectif réduit à disposition et à la faible représentation des diamètres supérieurs à 40 cm. De façon générale, l'état du hêtre est meilleur que celui du sapin et de l'épicéa.

A ce jour, de nombreuses études confirment le rôle prépondérant du diamètre pour les essences courantes, le pin étant une exception constante (chez ce dernier, aucune dépendance de l'état de santé en fonction du diamètre n'a pu être mise en évidence, cf. SCHLAEPFER *et al.*, 1985 ; EICHENBERGER, 1985 ; MANDALLAZ, 1985).

6.2. Position sociale

De façon générale, le rôle de ce facteur est complexe et changeant, ce qui est d'autant plus surprenant que les effets sont généralement très significatifs. A Sainte-Croix, l'état de santé du sapin est, toutes choses égales par ailleurs, fortement lié à la position sociale : à meilleure position, plus grande probabilité d'un état « sain ». D'autres études sembleraient indiquer que, lorsqu'existe une dépendance significative, elle aille dans le même sens, sauf pour le hêtre où l'inverse semble se produire. Relevons que la notion de position sociale peut être problématique dans les peuplements irréguliers (comme c'est partiellement le cas à Sainte-Croix). Les difficultés d'interprétation suggèrent également des interactions entre la position sociale, la longueur du houppier et le diamètre, qui résistent encore à l'analyse statistique.

La figure 5 illustre l'influence combinée du diamètre et de la position sociale chez le sapin (Sainte-Croix). Les autres paramètres ont été fixés arbitrairement à la moyenne observée dans les 254 individus. Une analyse des contrastes révèle que les

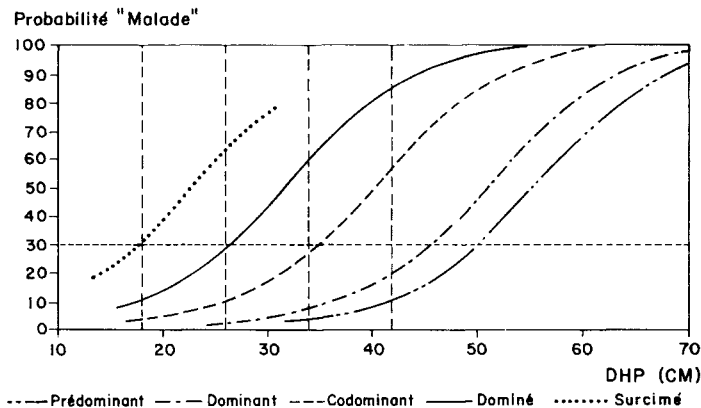


FIG. 5

Relation, chez le sapin, entre la probabilité ajustée d'être malade (en p. 100), le DHP et la position sociale (les autres facteurs étant fixés à leur valeur moyenne).

Relation, in fir, between adjusted probability of non healthy state (in p. 100), DBH and crown class (the other factors are kept fixed at their mean values).

courbes diffèrent significativement, sauf pour les arbres prédominants et dominants, qui peuvent être regroupés. Les lignes verticales se situent au dhp moyen des 4 positions sociales : par ordre croissant de gauche à droite « surcimés » « dominés » « codominants » « dominants/prédominants ». La ligne horizontale à 30 p. 100 correspond au pourcentage moyen d'arbres malades.

Le tableau 4 compare les fréquences observées avec les « estimations » données par la figure, qui n'est autre qu'un modèle logistique très simplifié. La concordance est bonne, sauf pour les codominants de dhp 30-40 cm, où en fait les houppiers courts sont sur-représentés.

Dans la littérature (SCHMID-HAAS, 1985), il est souvent mentionné que les arbres dominants sont plus atteints. En fait, il semblerait que ceci soit plutôt le fait de leur diamètre, en général plus grand, que de leur position proprement dite.

TABLEAU 4

*Comparaison qualitative entre courbes logistiques
et fréquences observées chez 254 sapins (Sainte-Croix).*

*Qualitative comparison between logistic curves and observed frequencies
in 254 firs (Sainte-Croix).*

(1) (2) (3)	(4)	Classes de diamètre		
		< = 30	30-40	> = 40
Dominants	0	5	65	
	0-5	5-15	15-95	
	3 (28)	7 (20)	75 (42)	
Codominants	0	47	77	
	0-15	15-50	50-90	
	5 (20)	30 (19)	80 (13)	
Dominés	12	50	80	
	5-45	45-80	80-95	
	15 (32)	60 (10)	92 (5)	
Surcimés	27	—	—	
	18-78	—	—	
	28 (65)	— (0)	— (0)	

(1) Pourcentage observé d'arbres malades.
Observed p. 100 of not healthy trees.

(2) Etendue couverte par la courbe logistique.
Range covered by logistic curve.

(3) Pourcentage estimé par la courbe au dhp médian de la classe.
Fitted p. 100 at median dbh of each crown class.

(4) Nombre d'arbres.

6.3. *Longueur du houppier*

De façon générale, il apparaît que l'influence du houppier, lorsqu'elle existe, soit favorable. Les arbres à houppier long ont une plus grande probabilité d'être sains que ceux à houppier court (ceci est le cas pour le sapin à Sainte-Croix).

6.4. *Degré de fermeture*

Ce paramètre est en général difficilement interprétable et diffère entre résineux et feuillus, sans révéler de tendance uniforme. Dans le cas de Sainte-Croix, les arbres en peuplement fermé ont une plus grande probabilité d'être sains. Le mauvais état des peuplements ouverts, auparavant très denses, est peut-être dû au déséquilibre causé par les interventions sylvicoles et par des phénomènes climatiques.

6.5. *Région*

Les facteurs régionaux (localisation, exposition) jouent un rôle important et doivent être ajustés pour l'étude des autres facteurs. A Sainte-Croix, les 3 régions définies *a priori* correspondent en fait à 3 zones bien délimitées quant à l'altitude, l'orientation et le relief : la région B est moyenne, A favorable et C défavorable. Dans ce cas, le découpage *a priori* correspond donc à une stratification *a posteriori* de l'état de santé.

6.6. *pH superficiel*

Sainte-Croix est pour l'instant notre seule étude avec mesure de pH superficiel (2 mesures à 10 cm par placette, moyenne utilisée comme facteur).

Une analyse plus fine que celle présentée dans le tableau 3 (les paramètres du pH sont estimés individuellement pour chacune des 3 essences) montre que la probabilité d'un état sain diminue significativement, chez l'épicéa et le sapin, quand le pH augmente, alors que la tendance est opposée mais non significative chez le hêtre (effectif trop faible pour une puissance satisfaisante). Ceci ne s'applique évidemment qu'à l'étendue observée du pH (4,5-7,0) et confirme l'expérience générale qu'un sol faiblement acide est propice à l'épicéa et au sapin, tandis qu'un milieu plus basique convient mieux au hêtre.

6.7. *Profondeur du sol*

Sainte-Croix est également notre seule étude avec mesure de profondeur (2 mesures par placette, moyenne utilisée comme facteur).

Nous constatons que plus le sol est profond, plus la probabilité d'un état **malade** est grande. Ce résultat surprenant est en accord, pour le sapin, avec les travaux de

SCHOEPFER & HRADEZKY (1985). Relevons cependant que le seuil de confiance atteint est faible étant donné le grand nombre de tests considérés, et nous nous garderons d'attribuer une trop grande importance à ce résultat.

7. Conclusions et perspectives

La régression logistique est une méthode très flexible et bien adaptée à l'étude des facteurs influençant l'état de santé; elle permet également une vérification légitime et pratique des modèles. Les ajustements sont encourageants pour de petits domaines et, pour de plus grands, doivent être améliorés par l'incorporation de variables exogènes décrivant les caractéristiques géologiques et atmosphériques (sans présumer de leur importance finale). La complexité du problème rend souvent l'interprétation délicate. Nous laisserons au sylviculteur le soin et la lourde responsabilité de tirer des enseignements pratiques des résultats présentés, en précisant que de nombreuses études restent à faire, tant au niveau de l'incorporation de variables exogènes et endogènes, qu'à celui du choix des stations et des méthodes statistiques employées. En ce qui concerne ce dernier point, nous réanalysons plusieurs études par la méthode multivariée des correspondances multiples. La comparaison entre le modèle logistique et l'analyse des correspondances sera l'objet d'une publication ultérieure.

Reçu le 5 décembre 1985.

Accepté le 24 mars 1986.

Remerciements

Nous tenons à remercier M. P. CHERBUIN, inspecteur forestier à Sainte-Croix, MM. H. EICHENBERGER et P. JOST, Office Cantonal des Forêts de Zürich, de leur précieuse collaboration.

Summary

Forest damage : attempt to model relationships

This paper presents a methodological approach to the analysis of relationships between the state of health (in spruce, fir and beech) and environmental as well as tree individual characteristics. Several forest damage inventories carried out in Switzerland have contributed to the development of the method. A detailed discussion of the results obtained in the district of Sainte-Croix (Jura) is given. The statistical analysis rests upon a logistic regression model which relates, for a given tree, the probability of an healthy state (the classification « healthy » - « not healthy » is made according to a threshold value of 20 p. 100 for the foliage loss) to several explanatory variables such as specie, diameter, crown length, crown class, region, ph, etc. Model validation procedures are emphasized, together with the distinction between exploratory and confirmatory analysis.

The probability for an healthy state is primarily a decreasing function of the diameter, but depends also, in a complex and often changing way, on crown class, crown length and other factors. The model can and should be extended with exogenous variables pertaining to air pollution.

Références bibliographiques

- COX D.R., 1980. *Analysis of Binary Data*. Chapman and Hall, London.
- COCHRAN W.G., 1977. *Sampling Techniques*, Wiley. New York.
- EICHENBERGER H., 1985. Ergebnisse der Waldschadeninventur 1984 im Kanton Zürich, Schweiz. *Z. Forstwesen*, **136** (11), 903-912.
- KENDALL M.G., STUART A., 1979. *The Advanced Theory of Statistics*. Vol. II, Chapter 30. **19**, Griffin, London.
- MANDALLAZ D., 1985. *Use of logistic regression models in the statistical analysis of forest damages*. Proceedings of the IUFRO S4.02, S6.05 Conference. Zürich, Switzerland. (A paraître).
- MCCULLAGH P., NELDER J.A., 1985. *Generalized Linear Models*. Chapman and Hall, London.
- PREGIBON D., 1981. *Logistic regression diagnostics*. *Annals of Mathematical Statistics*, 705-729.
- SAS User's guide, 1982. Sas Institute Inc.
- SCHLAEPFER R., MANDALLAZ D. *et al.*, 1985. Der Gesundheitszustand des Waldes im Revier Schaffhausen. Schweiz. *Z. Forstwesen*, **136** (1), 1-18.
- SCHLAEPFER R., 1985. *Relationship between diameter growth and crown parameters, in particular foliage loss, of Picea abies and Abies alba : result of a pilot survey in Switzerland*. Proceedings of the IUFRO S1.06-02 workshop. Tsukuba, Japan. A paraître.
- SCHMID-HAAS P., 1985. Der Gesundheitszustand des Schweizer Waldes. Schweiz. *Z. Forstwesen*, **136** (4), 251-273.
- SCHOEPPER W., HRADETZKY J., 1984. *Forstw. Cbl*, **103**, 231-248.
- WILLIAMS D.A., 1983. *The use of the deviance to test the goodness of fit of a logistic-linear model to binary data*. The GLIM newsletter, issue no. 6, 60-62.