

STATISTIQUE ET ANALYSE DES DONNÉES

FRANÇOIS HOULLIER

Comparaison de courbes et de modèles de croissance. Choix d'une distance entre individus

Statistique et analyse des données, tome 12, n° 3 (1987), p. 17-36.

http://www.numdam.org/item?id=SAD_1987__12_3_17_0

© Association pour la statistique et ses utilisations, 1987, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

Statistique et Analyse des Données

1987 - vol. 12 n° 3 p. 17-36

COMPARAISON DE COURBES ET DE MODELES DE CROISSANCE
CHOIX D'UNE DISTANCE ENTRE INDIVIDUS

François HOULLIER *

Inventaire Forestier National
Antenne de Recherches - Cellule Evaluation de la Ressource
Place des Arcades BP 1
34970 - Maurin / LATTES

Une large partie de ce travail a été réalisée au Laboratoire de Biométrie, U.A. 243 du C.N.R.S.,
Université Claude Bernard (Lyon I).

Résumé - *Le problème de la comparaison et de la typologie de courbes de croissance univariées est abordé : (1) en analysant la notion de distance entre courbes dans les différents espaces engendrés par les données expérimentales, les courbes sous-jacentes et les paramètres estimés par ajustement à un modèle; (2) puis en réalisant des analyses en composantes principales dans ces espaces. On étudie en particulier le cas où l'on s'intéresse aux paramètres estimés de courbes modélisées : on introduit alors une métrique spécialement conçue à cet effet et dite "métrique de sensibilité". Ces développements sont illustrés sur des courbes de croissance de peupliers.*

Summary - *This paper deals with methods related to the comparison of growth curves. The first part consists in defining different spaces generated either by the experimental data, the underlying growth curves, or the parameters estimated by fitting the data to a growth model. The relationships of the distances in these spaces are then elucidated. It is finally proposed to realize principal component analysis. A special metrics dedicated to the comparison of the estimated parameters of growth models is introduced : this metrics depends on the mathematical form of the model and is called "sensitivity metrics". Finally, an example concerning poplar girth growth is provided.*

Mots-clés : *modélisation, courbes de croissance, analyse en composantes principales, métrique, analyse de sensibilité.*

Indices de classification SMTA : 06-000, 06-070, 07-140

Manuscrit reçu le 15 novembre 1986

Révisé le 23 octobre 1987

0 - INTRODUCTION

L'analyse de la variabilité interindividuelle d'une famille de courbes de croissance expérimentales, longitudinales et univariées, c'est-à-dire d'un ensemble de mesures successives d'une même variable sur plusieurs individus, vise trois objectifs complémentaires : définir des groupes d'individus ayant des courbes de croissance similaires, interpréter les éléments qui différencient ces courbes et résumer les courbes en les réduisant à quelques paramètres ou variables synthétiques. ESTEVE et SCHIFFLERS [1] se sont placés dans le cas particulier où les mesures sont simultanées pour tous les individus et ont alors distingué trois stratégies pour parvenir à ces objectifs :

- La première consiste à procéder à une analyse factorielle des données initiales, éventuellement transformées, par exemple en remplaçant les données par les accroissements (cf [2], [3], [4] et [5, p.101-104]). L'analyse en composantes principales (ACP) permet ainsi à la fois de décomposer la variabilité des courbes en considérant la projection des vecteurs dates (chaque date de mesure détermine une variable) sur les composantes principales, de résumer la dispersion des courbes par projection sur les axes principaux du nuage des individus et de proposer une reconstitution des courbes de croissance en utilisant les coordonnées factorielles des individus.

- La deuxième stratégie consiste à procéder à un ajustement polynomial de chaque chronique individuelle, puis à comparer les coefficients obtenus pour les différents individus par analyse de variance multidimensionnelle.

- La troisième en est voisine puisqu'il s'agit d'ajuster un modèle non linéaire à chaque chronique, puis de comparer les paramètres de ces modèles par des méthodes inférentielles [6] ou descriptives (cf [7], [8], [9], [10] et [5, p.104-112]).

La distinction entre ces deux dernières approches repose sur deux éléments : d'une part, les modèles non linéaires ont souvent pour ambition de posséder des paramètres interprétables du point de vue biologique ou physique ([11], [12] et [13]), tandis que les modèles polynomiaux sont presque toujours exclusivement empiriques; d'autre part, le caractère linéaire de ces derniers autorise l'emploi de méthodes statistiques inférentielles pour lesquelles on ne dispose que de résultats asymptotiques ou approximatifs dans le cas des modèles non linéaires [14].

Dans leur revue, ESTEVE et SCHIFFLERS ont finalement affirmé leur préférence pour les deux premières stratégies, l'une parce que son pouvoir descriptif est supérieur, l'autre parce qu'elle permet de faire de l'inférence rigoureuse. On remarquera cependant que :

(1) l'ajustement de modèles linéaires, en particulier de modèles polynomiaux, ou non linéaires relève d'une même approche puisqu'il s'agit de procéder en deux phases : d'abord de réduire les courbes à quelques paramètres supposés pertinents, puis d'analyser la variabilité de ces paramètres; (2) il arrive que les différents individus soient mesurés un nombre variable de fois et à des dates quelconques; on est alors en présence de données manquantes; (3) lorsqu'on adopte un point de vue descriptif, le problème du choix d'une distance entre individus se pose aussi bien dans la première que dans la troisième démarche.

C'est autour de ces remarques que s'organise cet article, qui vise à proposer une synthèse des trois stratégies précitées dans le cadre d'une approche descriptive. La réflexion sur le choix d'une distance pertinente y joue un rôle central et peut être éclairé par quelques questions complémentaires :

- Lorsqu'on réalise directement une ACP sur les données expérimentales, plusieurs métriques peuvent être utilisées : la métrique identité (notée $I_{p,p}$, où p est le nombre de dates de mesures), la métrique diagonale de pondération par les variances à chaque date ou des métriques non diagonales tenant compte de l'autocorrélation présumée des mesures successives [3]. Peut-on justifier le choix de l'une ou l'autre de ces métriques en considérant que les données ne sont que la discrétisation d'un phénomène continu sous-jacent ?

- Lorsqu'on réalise un ajustement préalable à un modèle et qu'on effectue une ACP sur les paramètres estimés, on peut se demander si les métriques couramment utilisées (métrique identité, métrique diagonale de pondération par les variances interindividuelles des paramètres, ou métrique de Mahalanobis ([8], [15])) ont une signification claire du point de vue de l'impact des paramètres sur la forme et l'amplitude des courbes de croissance modélisées.

- Quels sont les rôles relatifs de la variabilité inter-individuelle et de la variabilité intra-individuelle (représentée sur les données brutes par l'autocorrélation des mesures successives, ou sur les paramètres par la métrique de Mahalanobis) dans les analyses ainsi réalisées ?

La démarche adoptée ici consiste : à formaliser les relations existant entre les différents espaces impliqués dans ce type d'analyses (§ 1); à préciser la notion de distance entre courbes dans ces espaces et à introduire une métrique particulière dans l'espace engendré par les paramètres d'un modèle de croissance (§ 2); à établir des liens exacts (lorsque c'est possible) ou approximatifs entre les analyses factorielles réalisées dans ces espaces (§ 3). Ces considérations sont illustrées par une étude de la croissance en circonférence du peuplier (§ 4).

1 - NOTATIONS ET DEFINITIONS

On étudie l'évolution d'une variable scalaire X , en fonction du temps t qui décrit un intervalle $]T_0, T_1[$ de \mathbb{R} , pour une famille de n individus notés $i : 1 \leq i \leq n$. Le terme de courbe de croissance correspond à plusieurs notions distinctes. Aussi va-t-on, dans un premier temps, préciser ses diverses acceptions en définissant plusieurs espaces distincts puis en analysant leurs interrelations.

1.1 - Courbes de croissance

On appelle courbe de croissance toute fonction $X(t)$ définie de $]T_0, T_1[$ dans \mathbb{R} . Dans la suite on suppose que $]T_0, T_1[$ est borné et on impose seulement à cette fonction d'appartenir à l'espace vectoriel des fonctions de carré intégrable $L^2(]T_0, T_1[)$, dont l'espace des fonctions continues de carré intégrable, $C^0(]T_0, T_1[)$, est un sous-espace vectoriel. On peut aussi se restreindre aux seules fonctions croissantes (mais cela ne change rien aux développements qui suivent). $(X_i(t))_{1 \leq i \leq n}$ désigne une famille de courbes de croissance de $L^2(]T_0, T_1[)$. Notons que toutes les courbes d'une même famille sont définies sur le même intervalle et qu'il s'agit là d'un préalable nécessaire à leur comparaison.

1.2 - Courbes de croissance expérimentales

On appelle courbe de croissance expérimentale de l'individu i le vecteur X_i^* de $\mathbb{R}^{p(i)}$, formé des mesures $X_i^*(t_j)$, réalisées aux dates successives t_j telles que $1 \leq j \leq p(i)$ et $T_0 \leq t_1 < t_2 < \dots < t_{p(i)} \leq T_1$. $(X_i^*)_{1 \leq i \leq n} = ((X_i^*(t_j))_{1 \leq j \leq p(i)})_{1 \leq i \leq n}$ définit une famille de courbes de croissance expérimentales. Remarquons que l'on ne fait pas *a priori* l'hypothèse de simultanéité des dates de mesure, ni *a fortiori* celles d'égalité des dimensions $p(i)$, ou d'équirépartition des dates de mesure sur $]T_0, T_1[$.

Lorsque ces dernières conditions sont vérifiées, on retrouve la situation envisagée par ESTEVE et SCHIFFLERS ([1], [3]) : $t_j = \bar{t} = T_0 + (j-1/2)(T_1-T_0)/p$ et $p(i) = p$. Une famille de courbes de croissance expérimentales peut alors être assimilée à un tableau quantitatif $X_{n \times p}^*$ d'éléments $X_j^* = X_i^*(t_j)$. On note alors \mathbb{R}^p l'espace engendré par les variables $X_j^* = X^*(\bar{t})$ et $E_p(]T_0, T_1[)$ le sous-espace vectoriel de $L^2(]T_0, T_1[)$ formé par les fonctions en escalier sur $]T_0, T_1[$ constantes sur chacun des p intervalles $]\bar{t}-1/2, \bar{t}+1/2[$.

1.3 - Modèles de croissance

On appelle modèle déterministe (générique) de croissance un sous-ensemble M_F de $L^2(JT_0, T_1[)$, formé des courbes $X(t)$ que l'on peut mettre sous la forme :

$$X(t) = F(t, Y) \quad (1)$$

F est une fonction définie de $JT_0, T_1[\times U$ dans \mathbb{R} , et U un domaine ouvert et convexe de \mathbb{R}^q .

Y est appelé le vecteur des paramètres de F . On ne s'intéresse ici qu'au cas, généralement vérifié dans la pratique, où $F(t, Y)$ est continument différentiable par rapport à Y sur $JT_0, T_1[\times U$. Chaque élément de M_F est appelé un modèle (particulier) de croissance. Dans la pratique, les modèles utilisés sont continus par rapport au temps si bien qu'on suppose dans la suite que M_F est inclus dans $C^0(JT_0, T_1[)$.

Dire que M_F est un bon modèle pour l'ensemble des individus étudiés revient à admettre que pour chaque individu i , on peut trouver au moins un vecteur Y_i dans U tel que $F(t, Y_i)$ "reconstitue bien" (au sens du § 1.4) la courbe $X_i(t)$. Cette hypothèse est évidemment essentielle et sa pertinence conditionne celle des analyses ultérieures. On se place ici dans le cas où cette supposition est valide. Une famille de modèles, $(F(t, Y_i))_{1 \leq i \leq n}$, peut alors être analysée dans deux espaces vectoriels distincts : dans $L^2(JT_0, T_1[)$ ou dans l'espace, \mathbb{R}^q , engendré par les paramètres.

1.4 - Identification d'un modèle de croissance

On note $G^j(t, Y)$ et on appelle fonction de sensibilité les dérivées partielles de F , supposées définies et continues, par rapport à chacun des paramètres Y^j :

$$G^j(t, Y) = (\partial F / \partial Y^j)_{t, Y} \quad (2)$$

Pour chaque individu i , échantillonné aux dates t_j , on définit la matrice de sensibilité H_i de dimension $(q, p(i))$, formée des éléments :

$$H_{j,k} = G^j(t_j, Y_i) \quad (3)$$

Les fonctions de sensibilité jouent un rôle crucial lors des différentes phases de la modélisation, en particulier lors de l'identification du modèle. Celle-ci consiste à estimer les paramètres Y_i tels que l'ajustement de la courbe modélisée $F(t, Y_i)$ à la courbe expérimentale $(X_i^*(t_j), 1 \leq j \leq p(i))$ soit optimal au sens d'un certain critère (exemple : critère du maximum

de vraisemblance) qui mesure la qualité de la reconstitution de la courbe expérimentale et qui est déterminé par le choix d'un modèle statistique pour l'erreur aléatoire ε_i [16] :

$$X_i^*(t) = F(t, Y_i) + \varepsilon_i \quad (4)$$

La caractérisation complète d'un modèle implique en fait la donnée conjointe de M_F et du modèle d'erreur, qui englobe aussi bien la variabilité propre de la croissance individuelle que la variabilité liée au procédé expérimental. La plupart des méthodes d'identification font appel à des algorithmes de minimisation qui requièrent l'existence et la continuité des fonctions de sensibilité [17], conditions que nous avons déjà supposées vérifiées.

On dit que le modèle est structurellement identifiable sur $]T_0, T_1[\times U$ si la fonction $F(t, \cdot)$ est bijective de \mathbb{R}^q dans M_F . Cette notion ne fait pas référence aux données expérimentales, mais aux seules propriétés mathématiques de F . On suppose dans la suite que cette condition est vérifiée. L'identifiabilité recouvre cependant une autre notion moins large : il s'agit, pour une chronique expérimentale donnée, X_i^* , de la possibilité de déterminer de façon unique un vecteur Y_i^* de U qui minimise le critère retenu. Cette notion met à la fois en jeu les propriétés mathématiques du modèle et la nature du plan d'échantillonnage temporel. Pour les modèles d'erreur usuels, on peut donner une condition nécessaire d'identifiabilité : "pour la chronique i ($1 \leq i \leq n$), il faut que : $|H_i^T H_i| \neq 0$ pour $Y = Y_i^*$ " [16, p.346]. (5)

On admettra dans la suite que cette condition est vérifiée et on fera de plus une hypothèse complémentaire de forme voisine :

$$\forall Y \in U, \text{ la matrice } Q_Y, \text{ d'élément } Q_Y^{kh} = \int_{T_0}^{T_1} G^k(t, Y) G^h(t, Y) dt, \text{ est définie ".} \quad (6)$$

En tout point Y , cette matrice est par ailleurs symétrique et positive si bien qu'elle détermine un produit scalaire sur \mathbb{R}^q .

Lorsque le modèle est identifiable, on peut à la fois estimer le vecteur Y_i^* et la variance de cet estimateur. Si le modèle est non linéaire, ce dernier résultat n'est cependant qu'asymptotique ([14], [18]). Si l'erreur aléatoire est homoscedastique et non autocorrélée, un estimateur de la variance de Y_i^* est :

$$\text{Var}^*[Y_i^*] = \sigma_i^2 (H_i^{*T} H_i^*)^{-1} \quad (7)$$

où : $\sigma_i^{*2} = S_i^* / (p(i)-q)$ est la variance estimée de ϵ_i , S_i^* la somme des carrés des écarts entre les valeurs expérimentales et les valeurs prédites $F(t_j, Y_i^*)$ et H_i^* la valeur estimée de H_i .

1.5 - Structure de M_F

Dans le cas général, M_F n'est pas un sous-espace vectoriel de $L^2(JT_0, T_1[[])$, mais seulement une variété définie par :

$$\begin{aligned} \Phi : U &\longrightarrow L^2(JT_0, T_1[[]) \quad (\text{où } \mathbb{R}^q \supset U) \\ Y &\longrightarrow F(t, Y) \end{aligned}$$

Les hypothèses faites précédemment assurent que Φ est un difféomorphisme de classe C^1 de U dans M_F et que M_F est une variété différentiable de dimension q [19]. Lorsque le modèle est linéaire en fonction des paramètres, $F(t, Y)$ s'écrit :

$$F(t, Y) = \sum_{1 \leq k \leq q} Y^k F^k(t) \quad \text{et} \quad G^k(t, Y) = F^k(t) \in C^0(JT_0, T_1[[]), \quad (8)$$

et Φ définit un isomorphisme de \mathbb{R}^q dans M_F qui constitue dans ce cas un sous-espace vectoriel de $L^2(JT_0, T_1[[])$.

2 - DISTANCES ENTRE INDIVIDUS

On cherche maintenant à munir chacun des espaces définis précédemment d'une métrique et à analyser les relations entre ces métriques. On considère dans la suite que l'espace des courbes de croissance constitue l'espace de référence, c'est-à-dire qu'on postule que l'analyse pertinente est celle réalisée dans $L^2(JT_0, T_1[[])$. La pertinence des analyses réalisées dans d'autres espaces sera donc jugée en fonction de la proximité de leurs résultats avec ceux de l'analyse similaire dans l'espace de référence. Comme les courbes de croissance sont en fait inaccessibles à l'expérimentateur dans leur intégralité, on cherche à munir les espaces \mathbb{R}^p , M_F et \mathbb{R}^q d'une distance ou d'une norme telle que les distances entre individus (c'est-à-dire entre courbes) soient conservées ou, pour le moins, bien approximées.

2.1 - Distance entre courbes

L'espace des courbes de croissance, $L^2([T_0, T_1])$, peut être muni du produit scalaire $(X|X')_{L^2} = \int X(t) X'(t) dt$ (où l'intégrale est prise entre T_0 et T_1) [20]. La norme associée dite " norme L^2 " permet de définir une distance entre deux courbes de croissance X_i et X_j :

$$d_{L^2}(X, X') = \|X - X'\|_{L^2} = \left(\int_{T_0}^{T_1} (X(t) - X'(t))^2 dt \right)^{1/2} \tag{9}$$

La distance associée à ce produit scalaire est la généralisation naturelle à une courbe de la distance euclidienne (dans \mathbb{R}) entre $X(t)$ et $X'(t)$ pour une date t fixée.

Remarquons que $L^2([T_0, T_1])$ peut être muni d'un autre produit scalaire, en considérant par exemple [20] :

$$(X|X') = \int_{T_0}^{T_1} \int_{T_0}^{T_1} X(t) A(t, t') X'(t') dt dt', \tag{10}$$

où A est un élément de $L^2([T_0, T_1] \times [T_0, T_1])$, tel que :

- $\forall (t, t') \in [T_0, T_1] \times [T_0, T_1], A(t, t') = A(t', t)$;
- et $(\iint X(t) A(t, t') X'(t') dt dt' = 0) \Rightarrow (X = 0 \text{ presque partout dans } [T_0, T_1])$.

2.2 - Distance entre courbes expérimentales

Le produit scalaire de $L^2([T_0, T_1])$ définit sur le sous-espace vectoriel $E_p([T_0, T_1])$ (cf § 1.2) une structure d'espace euclidien et la distance entre deux courbes expérimentales échantillonnées régulièrement aux mêmes dates est proportionnelle à :

$$d_E(X^*, X^{**}) = \left(\sum_{1 \leq j \leq p} (X^*(t_j) - X^{**}(t_j))^2 \right)^{1/2} \tag{11}$$

Cette distance coïncide avec la distance euclidienne usuelle dans \mathbb{R}^p . Si on considère deux courbes continues X et X' de $L^2([T_0, T_1])$ et qu'on les échantillonne régulièrement entre $[T_0, T_1]$, la différence $d_{L^2}(X, X') - d_E(X^*, X^{**})$ correspond à l'écart entre une intégrale et une somme de Riemann : elle est, de ce fait, négligeable dès que le nombre p d'observations est suffisant. Lorsque les courbes sont régulièrement échantillonnées sur $[T_0, T_1]$, le choix de la norme L^2 dans l'espace des courbes de croissance conduit donc à celui de la métrique identité dans l'espace des courbes expérimentales \mathbb{R}^p . Si on choisit une autre norme ou si le pas d'échantillonnage temporel n'est pas régulier, on aboutit en revanche à d'autres métriques dans \mathbb{R}^p .

2.3 - Distance entre modèles de croissance

Modèles linéaires

Lorsque le modèle est linéaire, M_F muni de la norme L^2 a une structure d'espace vectoriel euclidien, le produit scalaire, $(\cdot, \cdot)_{L^2}$, défini dans M_F se transporte exactement dans \mathbb{R}^q , où il est noté $(\cdot, \cdot)_{\mathbb{R}^q}$, et les distances entre individus sont conservées; soit $X(t) = F(t, Y)$ et $X'(t) = F(t, Y')$ deux éléments de M_F :

$$X(t) = \sum_{1 \leq k \leq q} Y^k G^k(t) \quad \text{et} \quad X'(t) = \sum_{1 \leq k \leq q} Y'^k G^k(t) \quad (\text{où } G^k(t, Y) = F^k(t))$$

$$(X|X')_{L^2} = \int_{T_0}^{T_1} \sum_{1 \leq k \leq q} \sum_{1 \leq h \leq q} Y^k Y'^h G^k(t) G^h(t) dt = \sum_{1 \leq k \leq q} \sum_{1 \leq h \leq q} Y^k Y'^h Q^{kh} = (Y|Y')_{\mathbb{R}^q}$$

et $d_{\mathbb{R}^q}(Y, Y') = \|Y - Y'\|_{\mathbb{R}^q} = d_{L^2}(F(t, Y), F(t, Y')) \quad (11)$

où Q est la matrice définie en (6) (pour un modèle linéaire, Q_Y est indépendante de Y , soit $Q = Q_Y$). Lorsque le modèle est linéaire, il est donc possible de définir une métrique euclidienne dans \mathbb{R}^q , telle que les distances entre courbes soient exactement conservées.

Modèles non linéaires

Quand le modèle est non linéaire, il y a deux possibilités : (1) on définit sur \mathbb{R}^q la distance non nécessairement euclidienne induite par $\Phi : d_{\mathbb{R}^q}(Y, Y') = d_{L^2}(F(t, Y), F(t, Y'))$; (2) on définit sur \mathbb{R}^q un produit scalaire $(Y|Y')_{\mathbb{R}^q}$ tel que la distance euclidienne associée, $d_{\mathbb{R}^q}(Y, Y')$, soit une "bonne" approximation de la distance $d_{L^2}(F(t, Y), F(t, Y'))$.

C'est cette seconde possibilité que l'on développe maintenant. En chaque point $X = F(t, Y)$ de M_F , on considère le plan tangent $T_F(Y)$ à la variété M_F . D'après (6), ce plan est un espace vectoriel de dimension q , isomorphe à \mathbb{R}^q , dont une base est formée par les fonctions de sensibilité $(G^k(t, Y), 1 \leq k \leq q)$. On munit alors $T_F(Y)$ de la métrique riemannienne qui coïncide avec la métrique définie dans $L^2(JT_0, T_1[)$ [19]. Soit $V(t)$ et $W(t)$ deux vecteurs de $T_F(Y)$:

$$\begin{aligned}
 V(t) &= \sum_{1 \leq k \leq q} v^k G^k(t, Y) \quad \text{et} \quad W(t) = \sum_{1 \leq k \leq q} w^k G^k(t, Y), \\
 (VIW)_{L^2} &= (VIW)_{T(Y)} = \sum_{1 \leq k \leq q} \sum_{1 \leq h \leq q} v^k w^h Q_Y^{kh}.
 \end{aligned}
 \tag{13}$$

On retrouve donc sur $T_F(Y)$ la métrique Q_Y précédemment introduite. La stratégie retenue consiste alors : (1) à se placer au point $X(t) = F(t, Y)$, où Y est la valeur moyenne des paramètres Y_i pour la famille des n modèles, puis (2) à approximer chaque modèle $F(t, Y_i)$ par une fonction du plan tangent $T_F(Y)$ (Figure 1) et (3) à analyser les modèles ainsi approximés dans $T_F(Y)$, muni de la base $(G^k(t, Y))$, $1 \leq k \leq q$ et de la métrique Q_Y , ce qui revient à analyser leurs paramètres dans \mathbb{R}^q muni de la base canonique et de cette même métrique.

Techniquement, cela conduit à supposer que les écarts $Y_i - Y$ sont petits et que la linéarisation du modèle par développement limité de $F(t, Y)$ en $Y_i - Y$ est valide :

$$F(t, Y_i) \approx F(t, Y) + \sum_{1 \leq k \leq q} (Y_i^k - Y^k) G^k(t, Y)
 \tag{14}$$

On doit cependant noter que : (1) l'approximation de $F(t, Y)$ dans le plan tangent $T_F(Y)$ n'est pas optimale au sens de la norme L^2 , car il ne s'agit pas d'une projection sur $T_F(Y)$, mais d'un développement limité; (2) la courbe $F(t, Y)$ appartient à M_F (du fait de la convexité supposée de U), mais cette courbe ne coïncide généralement pas avec la moyenne point par point des courbes $F(t, Y_i)$.

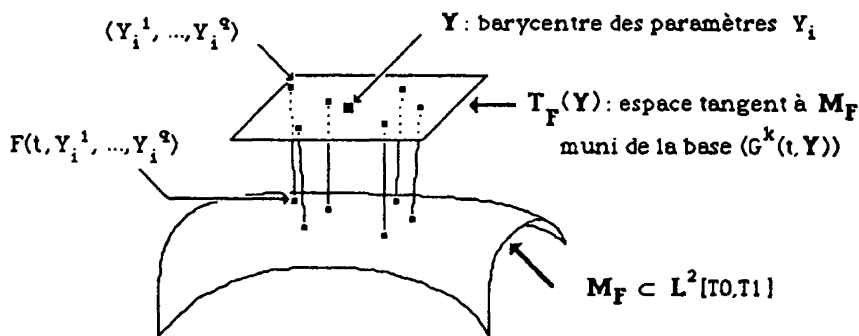


Figure 1 - Représentation schématique de la variété M_F dans $L^2([T0, T1])$ et de son plan tangent $T_F(Y)$ au point Y , barycentre des paramètres individuels Y_i ($1 \leq i \leq n$).

Soient i et i' deux individus, la distance $d_{L^2}(F(t, Y_i), F(t, Y_{i'}))$, est approchée par la distance $d_{\mathbb{R}^q}(Y_i, Y_{i'})$:

$$d_{L^2}(F(t, Y_i), F(t, Y_{i'})) \approx \left(\int_{T_0}^{T_1} \sum_{1 \leq k \leq q} \sum_{1 \leq h \leq q} (Y_i^k - Y_{i'}^k) G^k(t, Y) G^h(t, Y) (Y_i^h - Y_{i'}^h) dt \right)^{1/2},$$

$$d_{L^2}(F(t, Y_i), F(t, Y_{i'})) \approx d_{\mathbb{R}^q}(Y_i, Y_{i'}) = \left(\sum_{1 \leq k \leq q} \sum_{1 \leq h \leq q} (Y_i^k - Y_{i'}^k) Q_Y^{kh} (Y_i^h - Y_{i'}^h) \right)^{1/2} \quad (15)$$

Dans le cas d'un modèle non linéaire, on construit ainsi dans \mathbb{R}^q une métrique euclidienne tenant compte de la forme du modèle et approximativement équivalente à celle définie dans $L^2([T_0, T_1])$. La validité de cette approximation est intimement liée à celle du développement limité de $F(t, Y)$ autour de Y .

2.4 - Métrique de Mahalanobis

Par ailleurs, on définit dans \mathbb{R}^q la métrique de Mahalanobis, M , par [15] :

$$M^{-1} = \left(\sum_{1 \leq i \leq n} (p(i)-q) \text{Var}^*[Y_i^*] \right) / \left(\sum_{1 \leq i \leq n} (p(i)-q) \right).$$

Lorsque $p(i) = p$ pour tout i , M se simplifie :

$$M^{-1} = \left(\sum_{1 \leq i \leq n} \text{Var}^*[Y_i^*] \right) / n = \left(\sum_{1 \leq i \leq n} \sigma_i^{*2} (H_i^{*T} H_i^*)^{-1} \right) / n.$$

Si on suppose de plus que les variances, $\text{Var}^*[Y_i^*]$, sont égales, que le modèle est linéaire (ou que les fonctions de sensibilité $G_i^{k*}(t)$ sont très peu différentes de $G^k(t, Y)$) et que l'échantillonnage temporel est régulier sur $[T_0, T_1]$, alors M est égal au produit $H^{*T} H^*$ (à un facteur multiplicatif près) et M représente une somme de Riemann de Q_Y (H^* est la matrice de sensibilité H prise au barycentre des paramètres estimés Y_i^* , noté Y^* ; cf (3)).

Cette métrique permet de prendre en compte la variance intra-individuelle moyenne des paramètres estimés, si bien qu'une ACP munie de M peut être considérée comme une forme d'analyse discriminante.

3 - ANALYSE EN COMPOSANTES PRINCIPALES SUR LES PARAMETRES

Maintenant qu'on a défini les différents espaces en présence et qu'on les a muni de métriques, le lien entre les analyses effectuées dans ces espaces est immédiat.

Quand les données sont équiréparties sur $]T_0, T_1[$, les développements précédents permettent de déduire du choix d'un produit scalaire dans $L^2(]T_0, T_1[)$ une métrique euclidienne dans l'espace engendré par les paramètres estimés, \mathbb{R}^p . On peut alors réaliser une ACP dans \mathbb{R}^p , muni de cette métrique.

Lorsqu'on s'intéresse à la croissance en tant que phénomène continu représenté par un modèle, on peut adopter deux points de vue :

- Le premier consiste à ne voir l'ajustement au modèle que comme une étape intermédiaire. La comparaison des courbes de croissance peut alors être réalisée au moyen d'une analyse factorielle sur le tableau de distances $(d_{L^2}(F(t, Y_i), F(t, Y_{i'})))_{1 \leq i \leq n, 1 \leq i' \leq n}$ [21]. On obtient ainsi une reconstitution euclidienne du nuage des individus, mais l'analyse n'est pas directement interprétable en termes de paramètres du modèle.

- Le second point de vue suppose que les paramètres ont une interprétation phénoménologique, voire biologique ou physique, qui intéresse l'expérimentateur. Il est alors possible de réaliser une ACP dans l'espace \mathbb{R}^q muni de la métrique Q_Y , selon le schéma de dualité proposé à la figure 2 [21].

Lorsque le modèle est linéaire, les résultats précédents assurent que : (1) l'analyse réalisée sur la matrice de distances $(d_{L^2}(F(t, Y_i), F(t, Y_{i'})))_{1 \leq i \leq n, 1 \leq i' \leq n}$ coïncide avec l'analyse réalisée dans \mathbb{R}^q muni de la métrique Q_Y ; (2) les analyses réalisées dans \mathbb{R}^q et \mathbb{R}^p (lorsque les chroniques sont échantillonnées simultanément et régulièrement) coïncident approximativement, l'écart étant dû aux effets de la discrétisation expérimentale. Lorsque le modèle n'est pas linéaire, ces résultats ne sont plus qu'approximatifs.

On note aussi que, sous certaines conditions (§ 2.4), les ACP effectuées dans \mathbb{R}^q muni, d'une part, de la métrique de Mahalanobis et, d'autre part, de Q_Y donnent des résultats voisins, bien que leur principe soit différent. Ce paradoxe apparent tient au fait que les fonctions de sensibilité interviennent pour linéariser le modèle aussi bien lors de l'identification (linéarisation au voisinage de chaque courbe) que lors du transport des métriques (linéarisation au voisinage de la courbe des paramètres moyens). L'écart entre ces deux ACP est lié à la non

linéarité du modèle, aux effets de la discrétisation expérimentale, aux irrégularités de l'échantillonnage temporel et/ou à la variabilité interindividuelle des variances résiduelles.

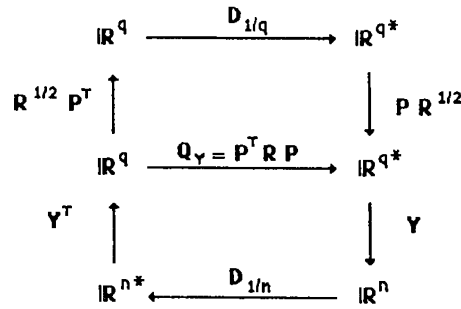


Figure 2 - Schéma de dualité de l'ACP sur les paramètres avec la métrique Q_Y : les facteurs sont obtenus par diagonalisation de $Y^{*T} D_{1/n} Y^* Q_Y$. (La partie supérieure du schéma a été ajoutée pour une raison technique, afin que la matrice à diagonaliser soit symétrique.)

4 - EXEMPLE

Les données étudiées concernent 32 courbes de croissance en circonférence moyenne provenant de plantations clonales de peuplier réalisées dans quatre localités et soumises à des traitements différents (fertilisation, travail du sol). Elles ont été présentées par BAILLY [9], qui a ajusté le "modèle de Nelder" [6] (par la méthode des moindres carrés ordinaires) à chaque chronique :

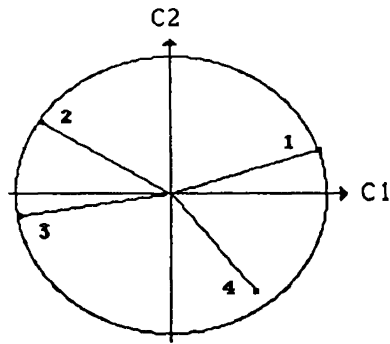
$$X(t) = Y^5 + Y^1 (1 - Y^4 \exp((Y^2 - t) / Y^3)) (-1/Y^4),$$

où t est l'âge de la plantation, Y^1 l'amplitude de la croissance, Y^2 la date du maximum de croissance, Y^3 une constante de temps qui détermine la vitesse de croissance, Y^4 un paramètre de forme qui fixe la valeur relative du maximum de croissance par rapport à l'amplitude et Y^5 la circonférence moyenne au moment de la plantation (les valeurs, Y^* , estimées par BAILLY sont données au tableau 1). Même si on peut en proposer des interprétations phénoménologiques ou biologiques ([9], [13]), le choix de ce modèle reste largement empirique.

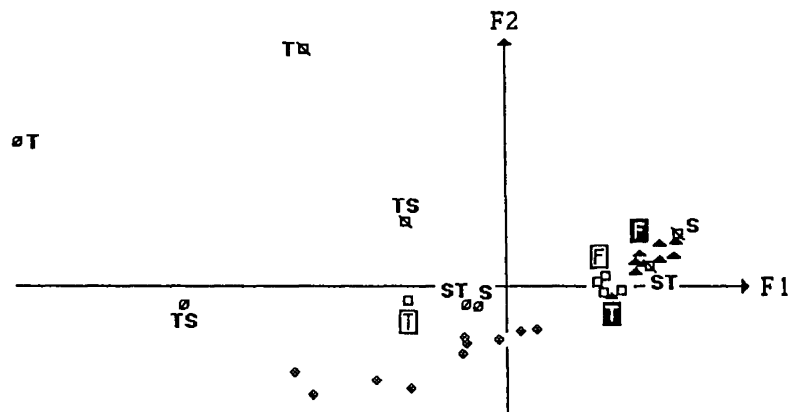
Essai	Clone	Traitement	Y ¹	Y ²	Y ³	Y ⁴	Y ⁵
1	I214	témoin	196,93	12,74	14,56	- 0,41	11,8
1	I214	témoin puis sarclage	138,05	5,22	9,62	- 0,84	10,0
1	I214	sarclage puis témoin	181,44	- 1,15	15,07	- 1,19	10,4
1	I214	sarclage	167,28	2,24	8,82	- 0,91	11,3
1	Robusta	témoin	140,15	12,70	14,84	- 0,43	9,9
1	Robusta	témoin puis sarclage	124,13	2,46	17,18	- 1,07	9,1
1	Robusta	sarclage puis témoin	139,99	0,96	12,64	- 1,03	10,7
1	Robusta	sarclage	124,21	2,42	8,56	- 0,88	9,6
2	I45/51	néant	108,77	- 2,64	12,81	- 1,32	11,8
2	Harvard	néant	111,66	0,24	7,20	- 1,15	12,8
2	Onda	néant	114,57	0,27	8,60	- 1,13	13,9
2	Lux	néant	158,97	1,49	18,44	- 1,12	11,0
2	I55/56	néant	95,45	0,08	11,75	- 1,04	9,4
2	Alcinde	néant	108,25	- 1,45	9,63	- 1,27	10,0
2	Robusta	néant	105,62	0,85	13,18	- 1,00	11,8
2	Scrotina	néant	137,59	- 0,28	13,78	- 1,07	10,6
2	Virginie	néant	95,46	- 0,15	9,68	- 1,09	8,3
2	I214	néant	166,37	- 1,90	17,73	- 1,16	10,9
3	I214	témoin	135,94	3,59	11,92	- 0,74	10,0
3	I214	fertilisé	151,52	5,15	7,37	- 0,44	10,2
3	I214	fertilisé	146,96	4,69	6,68	- 0,45	9,9
3	I214	fertilisé	145,16	4,88	6,81	- 0,43	10,0
3	I214	fertilisé	147,88	4,35	7,60	- 0,54	10,0
4	I214	témoin	136,56	2,60	6,47	- 0,77	12,1
4	I214	fertilisé	164,89	1,57	9,52	- 0,98	12,1
4	I214	fertilisé	166,98	0,87	10,20	- 1,01	11,8
4	I214	fertilisé	150,29	1,79	7,71	- 0,94	11,8
4	I214	fertilisé	161,63	0,63	9,86	- 1,06	11,8
4	I214	fertilisé	151,65	1,80	8,23	- 0,94	12,0
4	I214	fertilisé	159,65	1,28	9,71	- 0,99	10,9
4	I214	fertilisé	190,68	0,19	14,49	- 1,08	12,9
4	I214	fertilisé	151,49	1,94	8,26	- 0,88	10,9

Tableau 1 - Croissance en circonférence du peuplier : valeur des paramètres estimés par ajustement au modèle de Nelder (Données communiquées par Alain BAILLY, A.FO.CEL). (Voir [9] pour les données originales et pour connaître les conditions d'acquisition des données. Dans les essais 3 et 4, les modalités de fertilisation varient selon les parcelles. Dans certains cas, l'ajustement global d'une chronique expérimentale à un modèle unique est discutable dans la mesure où l'expérimentateur a sciemment modifié l'environnement au cours du temps; exemple : sarclage jusqu'à 5 ans, puis arrêt du sarclage).

L'ACP décrite à la figure 2 a été programmée (calcul de Q_Y par intégration numérique, puis ACP proprement dite), puis appliquée au tableau $Y^*_{32 \times 4}$, muni de la norme Q_Y et sur l'intervalle]0,25[ans, qui correspond à une durée maximale de révolution pour les peupliers cultivés. Le cinquième paramètre, Y^5 (circonférence moyenne initiale des arbres), n'a pas été pris en compte dans l'analyse.



3.a) Cercle des corrélations (les numéros représentent les paramètres) :
les corrélations sont définies au sens de la métrique Q_Y



premier essai : clone I214 = \square clone Robusta = σ
 T = témoin TS= témoin puis sarclage ST= sarclage puis témoin S= sarclage

deuxième essai : comparaison de clones = \diamond

troisième essai : clone I214 = \square \square = témoin \square = fertilisé

quatrième essai : clone I214 = \blacktriangle \blacksquare = témoin \blacksquare = fertilisé

3.b) Représentation factorielle du nuage des courbes de croissance dans le plan (F1, F2)

Figure 3 - Etude de la variabilité de la croissance en circonférence du peuplier :
ACP sur les paramètres estimés avec la métrique Q_Y .

Les résultats sont donnés à la figure 3 : (1) on distingue un "effet taille" (opposition des individus à croissance rapide et à asymptote élevée, aux individus à croissance lente et à asymptote basse) et un effet "forme et position de la courbe sur l'axe des temps" (opposition des individus à croissance précoce aux individus à croissance tardive); (2) ils mettent en évidence l'effet des localités, celui des traitements (fertilisation, travail du sol) à l'intérieur des localités ainsi que la vigueur plus ou moins grande des clones; (3) le fait que les deux premiers axes absorbent respectivement 83 et 16% de l'inertie totale du nuage suggère qu'une reparamétrisation du modèle serait certainement judicieuse dans le contexte particulier du jeu de données étudié.

Il n'était pas possible de réaliser une ACP sur les données initiales car les âges de mesure ne coïncidaient pas tous. Cependant, une comparaison a été menée en simulant de manière purement déterministe des données à partir des modèles ajustés. Les ACP réalisées sur les données simulées (avec la métrique $I_{p,p}$) et sur les paramètres (avec la métrique Q_Y) se sont avérées très voisines tant du point de vue de l'inertie absorbée par les axes principaux que du point de vue des représentations factorielles.

Des essais effectués sur d'autres données ont cependant indiqué que, lorsque les paramètres sont mal estimés, les deux ACP (sur données réelles et sur paramètres) peuvent donner des résultats assez différents; les individus dont la représentation factorielle est "la moins stable" (ou "la moins bien conservée" d'une analyse à l'autre) étant ceux pour lesquels l'ajustement est le plus mauvais (BAILLY, comm. pers.).

5 - DISCUSSION

Une première remarque concerne le rôle privilégié conféré à l'espace des courbes de croissance. Ce choix revient à admettre que l'intérêt de l'expérimentateur se porte sur le phénomène sous-jacent, supposé continu, et qu'il est capable de définir une distance entre courbes de croissance. Deux stratégies se sont finalement dégagées : l'une vise à reconstituer la variabilité des courbes en se limitant aux données mesurées, c'est-à-dire en considérant des courbes discrétisées; l'autre vise à privilégier le caractère continu de la croissance (caractère parfois artificiel; voir l'exemple des peupliers), mais elle requiert des hypothèses supplémentaires sous la forme d'une classe particulière de courbes de croissance, c'est-à-dire d'un modèle. Il est alors évident que le choix de ce modèle est crucial pour la suite de l'analyse. Même si des outils existent, qui permettent de construire de tels modèles ([11] à [13]), puis de les identifier ([14] à [18]), ce choix reste généralement subjectif et doit être

considéré avec prudence ([13], [22]). Dans ce cadre, la distinction entre modèles linéaires (par exemple, polynomiaux) et non linéaires ne se justifie que par des considérations techniques : les résultats obtenus sont exacts pour des modèles linéaires et approximatifs pour des modèles non linéaires (§ 2.3).

L'utilisation d'un modèle, préalablement à l'analyse de la variabilité interindividuelle des courbes de croissance, offre d'autres avantages. Elle permet de se libérer partiellement de certaines contraintes portant sur le plan d'échantillonnage temporel (simultanéité et régularité des mesures). Elle met ainsi en évidence la nécessité de la définition d'une période d'étude commune par l'utilisateur. On remarque alors le rôle crucial joué par le choix *a priori* de l'intervalle]T0,T1[; en changer, c'est changer la question posée par l'expérimentateur à l'analyse et cela se traduit par une modification de la métrique Q_Y dans \mathbb{R}^q et, en corollaire, par une modification des résultats de l'ACP. L'utilisation d'un modèle permet enfin de réaliser une première réduction des données qui facilite le dépouillement de l'analyse lorsque les paramètres ont une interprétation phénoménologique, voire physique ou biologique, claire.

On a privilégié ici un point de vue déterministe en insistant sur la variabilité interindividuelle des courbes. Il s'est cependant avéré nécessaire d'élargir ce point de vue en introduisant un modèle d'erreur pour l'identification du modèle. Cet aspect probabiliste est ensuite réapparu au travers de la métrique de Mahalanobis. On peut cependant imaginer d'autres pistes :

- L'une, théorique, consiste à définir explicitement une courbe de croissance comme la superposition d'un modèle déterministe et d'un processus aléatoire qui rend compte de la variabilité propre de la croissance et des erreurs d'expérimentation (cf (4)). Il est alors possible de choisir un produit scalaire dans $L^2(JT0,T1[)$ qui tienne compte de la nature du processus stochastique (cf (10)). La démarche adoptée dans cet article peut ainsi se généraliser à cette situation.

- l'autre, plus pratique, consiste à traduire graphiquement la matrice de variance des paramètres par un ellipsoïde de confiance (dont on sait qu'il est lié à des hypothèses de normalité parfois douteuses et qu'il n'a qu'une valeur indicative dans le cas des modèles non linéaires) puis à représenter la projection de cet ellipsoïde sur les axes ou sur les plans factoriels.

Pour des modèles non linéaires, il est clair que la pertinence de la métrique Q_Y dépend de la validité du développement limité de $F(t,Y)$ autour de Y : c'est-à-dire de l'intensité de la non linéarité de F et de la dispersion des paramètres Y_i autour de Y . Pour apprécier l'effet

de cette non linéarité, on peut envisager d'associer à chaque point Y de U (dans l'espace des paramètres) la matrice Q_Y , qui définit une métrique euclidienne (si la condition (6) est vérifiée), puis d'étudier les différences entre les matrices Q_{Y_i} et Q_Y .

Une telle approche amène à considérer une autre possibilité. Le choix de la métrique Q_Y a été guidé ici par la recherche d'une métrique euclidienne conservant "au mieux" les distances définies dans l'espace des courbes de croissance : cette approche a conduit à privilégier le point Y , barycentre des individus et à analyser les courbes individuelles par référence à la courbe $F(t, Y)$ (qui ne coïncide pas avec la courbe moyenne si le modèle est non linéaire). Une autre démarche consiste à ne pas définir une telle courbe de référence, mais à chercher directement un compromis Q entre les différentes métriques locales, Q_{Y_i} . Formellement, on peut définir Q de la même façon que dans la méthode Statis ([5], [23]) : (1) en considérant les Q_{Y_i} comme des opérateurs plongés dans un espace euclidien (le produit scalaire étant la trace); (2) puis en leur appliquant la technique du positionnement multidimensionnel; (3) et en définissant Q comme une combinaison linéaire des opérateurs Q_{Y_i} (les coefficients de la combinaison étant les coordonnées, toutes positives ou nulles, du premier vecteur propre unitaire de la matrice des produits scalaires des opérateurs Q_{Y_i} et Q_{Y_j}). Cette méthode a l'avantage de permettre une analyse immédiate de l'éloignement relatif des métriques Q_{Y_i} (éloignement dû à la non linéarité du modèle; pour un modèle linéaire, il est en effet aisé de vérifier que Q et Q_Y coïncident, ainsi que toutes les matrices Q_{Y_i}). La signification concrète de Q n'est cependant pas évidente dans le cas des courbes de croissance.

La méthode développée ici et les résultats obtenus ne sont en fait pas limités aux seules courbes de croissance. Ils se généralisent à tout modèle défini par une fonction $F(s, Y)$ où s est un élément d'un pavé ouvert de \mathbb{R}^r et Y appartient à un ouvert convexe de \mathbb{R}^q . Les intégrales simples sont alors remplacées par des intégrales multiples. Cette méthode s'applique aussi aux modèles différentiels pour lesquels il n'existe pas nécessairement de solution analytique :

$$dX/dt = f(t, Y)$$

$$X(t_0) = x_0,$$

La seule difficulté rencontrée est alors d'ordre numérique puisque les fonctions de sensibilité de X (relativement à Y) ne sont pas explicites et s'obtiennent par intégration numérique [12].

Il faut enfin revenir sur le rôle central joué par les fonctions de sensibilité du modèle. Ce sont en effet elles qui interviennent dans la linéarisation du modèle et qui permettent ainsi d'établir un lien entre deux champs souvent disjoints des mathématiques appliquées : la modélisation et l'analyse des données. C'est pourquoi l'expression "métrique de sensibilité" a été proposée pour désigner Q_Y ([5], [10]).

BIBLIOGRAPHIE

- [1] ESTEVE, J. et SCHIFFLERS, E., "Discussion et illustration de quelques méthodes d'analyse longitudinale", in *Proceedings of the 9th international biometric conference*, Biometric Society (Boston), 1976, pp. 463-480.
- [2] BIROT, Y. et JOANNES, H., "Variabilité juvénile de la phénologie et des allures de croissance chez le Douglas", IUFRO (Göttingen), 1973, 14 pages.
- [3] ESTEVE, J. et SCHIFFLERS, E., "Some aspects of growth in laboratory mice : statistical analysis of a family of growth curves", *Ann. Zool. Ecol. anim.*, 1974, vol. 6, n° 3, pp. 449-461.
- [4] LEMOINE, B., "Application de l'analyse factorielle à l'étude de la croissance en hauteur des arbres : exemple du Pin maritime", *Ann. Sci. forest.*, 1981, vol. 38, n° 1, pp. 31-54.
- [5] PERNIN, M.O., *Contribution à la méthodologie d'analyse des données longitudinales : exemple de la croissance humaine*, Thèse de doctorat, Univ. Cl. Bernard (Lyon I), 1986.
- [6] DEBOUCHE, C., *Application de la régression non linéaire à l'étude et à la comparaison de courbes de croissance longitudinales*, thèse, Faculté des Sciences agronomiques de Gembloux, 1977.
- [7] CHASSE, J.L. et LEGAY, J.M., "Etude des courbes de croissance de quelques communes de la communauté urbaine de Lyon de 1876 à 1975", *J. Soc. stat. Paris*, 1980, tome 121, n° 3, pp. 143-152.
- [8] PAVE, A., CORMAN, A. et BOBILLIER-MONOT, R., "Utilisation et interprétation du modèle de Gompertz : application à la croissance de jeunes rats musqués", *Biom. Prax.*, vol. 26, pp. 123-140, (1986).
- [9] BAILLY, A., *Modélisation de la croissance en circonférence du peuplier : premiers résultats*. AFOCEL 1985, *Annales de recherches sylvicoles*, 1986, pp. 359-397.
- [10] HOULLIER, F., *Echantillonnage et modélisation de la dynamique des peuplements forestiers : application au cas de l'Inventaire Forestier National*, Thèse de doctorat, Univ. Cl. Bernard (Lyon I), 1986.
- [11] BERTALANFFY, L., *Théorie générale des systèmes*, Dunod (Paris), 1973.

- [12] PAVE, A., Contribution à la théorie et à la pratique des modèles mathématiques pour l'analyse quantitative des systèmes biologiques, Thèse de doctorat d'état, Univ. Cl. Bernard (Lyon I), 1980.
- [13] HOULLIER, F., Construction et interprétation de modèles dynamiques : exemples forestiers, in les Cahiers d'EDORA 1, INRIA (Sophia-Antipolis, à paraître).
- [14] MESSEAN, A., Application de la géométrie différentielle à la statistique du modèle non linéaire, Thèse de Docteur-Ingénieur, Univ. Paris-Sud (Orsay), 1984.
- [15] DAGNELIE, P., Analyse statistique à plusieurs variables, Presses agronomiques (Gembloux), 1975.
- [16] BECK, J.V. et ARNOLD, K.J., Parameter estimation in engineering and science, Wiley & sons (New York), 1977.
- [17] BARD, Y., Non linear parameter estimation, Academic Press (New York), 1974.
- [18] GALLANT, R., "Testing a subset of the parameters of a nonlinear regression model", J. am. stat. Ass., 1975, vol. 370, n° 352.
- [19] DOUBROVINE, B., NOVIKOV, S., FOMENKO, A., Géométrie contemporaine, méthodes et applications, deuxième partie : géométrie et topologie des variétés, Editions MIR (Moscou), 1982.
- [20] TRENOGUINE, V., Analyse fonctionnelle, Editions MIR (Moscou), 1985.
- I [21] CAILLIÉZ, F. et PAGES, J.P., Introduction à l'analyse des données, SMASH (Paris), 1976.
- [22] FELLER, W., "On the logistic law of growth and its empirical verifications in biology", (1940), in Applicable mathematics of non physical phenomena, OLIVEIRA-PINTO, F. et CONNOLLY, B.W. (éditeurs), John Wiley, 1982, pp. 125-138.
- [23] ESCOUFIER, Y., "L'analyse conjointe de plusieurs matrices de données", in Biométrie et temps, JOLIVET, E. et al (éditeurs), 1980, pp. 59-76.

Remerciements - Je remercie vivement Alain Bailly, Alain Franc, Alain Pavé et Marie-Odile Pernin, avec qui j'ai collaboré sur ce sujet, ainsi que les deux rapporteurs pour leurs remarques.