

# REVUE DE STATISTIQUE APPLIQUÉE

S. CHAMPELY

CH. DELEUZE

## **Co-inertie de deux variables aléatoires hilbertiennes et approximation B-spline application en foresterie**

*Revue de statistique appliquée*, tome 43, n° 1 (1995), p. 91-107

<[http://www.numdam.org/item?id=RSA\\_1995\\_\\_43\\_1\\_91\\_0](http://www.numdam.org/item?id=RSA_1995__43_1_91_0)>

© Société française de statistique, 1995, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). 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/>

## **CO-INERTIE DE DEUX VARIABLES ALÉATOIRES HILBERTIENNES ET APPROXIMATION B-SPLINE APPLICATION EN FORESTERIE**

S. Champely, Ch. Deleuze

*Biométrie, Génétique et Biologie des Populations – URA 243, Université Lyon I,  
43 bd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France.*

### **RÉSUMÉ**

L'Analyse en Composantes Principales (ACP) d'une variable aléatoire hilbertienne est un modèle théorique pour définir l'ACP de courbes. Elle nécessite une approximation, ici B-spline, et une estimation de l'opérateur de covariance pour être réalisable. En présence de deux collections appariées de courbes dont on souhaite étudier les variations conjointes, nous proposons un modèle théorique de co-inertie (CHESSEL et MERCIER, 1993) de deux variables aléatoires hilbertiennes, extension de l'analyse inter-battery (TUCKER, 1958). Il faut alors une double approximation par les B-splines dans les deux espaces de courbes et une estimation statistique de l'opérateur de covariance croisée pour la pratiquer.

La nature des données conduit parfois à se limiter à des courbes positives ou monotones (croissance par exemple). Des bases particulières de splines couplées à certaines contraintes sur les coefficients permettent de respecter la «nature» des courbes estimées. Le retour à une formulation B-spline reste possible.

L'application concerne l'étude simultanée de la croissance en hauteur et en circonférence d'une centaine d'Épicéas communs (*Picea abies*).

**Mots-clés :** *analyse en composantes principales, co-inertie, courbes de croissance, foresterie, processus.*

### **SUMMARY**

Principal Components Analysis (PCA) of a random variable with values in a separable Hilbert space is a theoretical model to define PCA of curves. It needs an approximation, here processed with B-spline functions, and an estimation of the covariance operator. We propose a theoretical model of co-inertia analysis (CHESSEL and MERCIER, 1993) for two hilbertian random variables for studying simultaneous variations in two matched sets of curves. This is a straightforward generalization of TUCKER's inter-battery analysis (1958). We explain the double approximation in each space by B-splines and the sample estimation of the crossed covariance operator.

The very nature of data leads sometimes to positive or monotonous curves (for instance growth curves). Special spline bases can be used with constraints on the parameters to impose

positivity or monotonicity. These curves can be put into the more convenient form of a B-spline basis.

The motivating example for the present paper is the simultaneous growth in height and girth of about a hundred Norway spruces (*Picea abies*).

**Keywords :** *co-inertia analysis, forestry, growth Curves, principal components analysis, process.*

## 1. Introduction

L'étude par des méthodes d'analyse multivariée linéaire d'un ensemble de courbes discrétisées peut être menée en considérant directement les données sous forme de fonctions. Cette démarche est classique en météorologie ou en traitement du signal, sous le nom d'analyse de Karhunen-Loève (LOÈVE, 1963).

Il n'en va pas de même en foresterie où l'on peut seulement citer la contribution d'HOULLIER (1987). L'auteur propose de réaliser une approximation de l'Analyse en Composantes Principales (ACP) de courbes définies par un modèle paramétrique, par l'intermédiaire d'une ACP sur les paramètres du modèle moyennant une métrique héritée d'un développement limité. Il souligne que dans le cas d'un modèle linéaire en ses paramètres, cette solution s'avère exacte.

Afin d'obtenir une méthode suffisamment générale d'analyse de courbes, il est bon de s'affranchir de la contrainte que représente l'établissement d'un modèle paramétrique unique adéquat, ce qui n'est pas toujours possible dans le cas de courbes chaotiques ou de formes très diversifiées. En outre, des considérations de biais qualitatifs (GASSER *et al.*, 1984), du fait que certaines caractéristiques ne peuvent être prises en compte de par la structure du modèle postulé, peuvent aussi incliner au choix d'un modèle semi-paramétrique. EL FAOUZI et ESCOUFIER (1991) proposent un modèle linéaire en ses paramètres et «souple» : l'approximation par les splines de régression que nous reprenons.

L'originalité de l'étude réside dans l'analyse conjointe de deux mesures fondamentales de l'arbre, la hauteur et la circonférence, par une méthode de couplage de fonctions appariées.

La section 2 décrit un modèle théorique d'analyse, englobant l'ACP de courbes. La suivante en propose un semblable pour le couplage de deux analyses. L'espace utilisé pour réaliser l'étude des courbes est ensuite rapidement présenté. La section 5 rappelle les propriétés essentielles de l'outil d'approximation, les fonctions B-splines, tandis que la suivante expose des modèles dérivés, adaptés à différentes contraintes naturelles des courbes (monotonie...). Les sections 7 et 8 proposent les approximations B-splines de l'ACP puis du couplage des courbes. L'application au modèle biologique fera l'objet de la dernière section.

## 2. Analyse en composantes principales d'une variable aléatoire hilbertienne

Soient  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $\mathcal{X}$  (supposée centrée, sans perte de généralité) une variable aléatoire d'ordre 2 définie sur cet espace à valeurs dans un espace de Hilbert réel séparable  $H$  dont le produit scalaire sera noté  $(\cdot|\cdot)_H$ .

Elle détermine une application linéaire continue de Hilbert-Schmidt  $u$ , de  $H$  dans  $L^2(\Omega, \mathcal{A}, P)$ , l'espace de Hilbert des variables aléatoires réelles de carré intégrable :

$$u : h \rightarrow (\mathcal{X}|h)_H, \quad (1)$$

et son adjoint  $u^\bullet$  :

$$u^\bullet : \mathcal{Z} \rightarrow E(\mathcal{X}Z). \quad (2)$$

L'ACP de  $\mathcal{X}$  est définie (DAUXOIS et POUSSE, 1976) comme la recherche de l'élément  $a$  de  $H$  de norme unité, appelé fonction orthogonale empirique, tel que  $u(a)$  soit de variance maximum, *et itérations sous contraintes d'orthonormalité dans  $H$ .*

Soit

$$v = u^\bullet u, \quad (3)$$

l'opérateur de  $H$  dit de covariance. Il est nucléaire auto-adjoint positif. Les fonctions orthogonales empiriques s'obtiennent par décomposition spectrale de  $v$  :

$$v = \sum_{\alpha} \lambda_{\alpha} (a_{\alpha} | \cdot)_H a_{\alpha} \quad (\text{théorème de Mercer}). \quad (4)$$

Les composantes principales :

$$C_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} u(a_{\alpha}) \quad (5)$$

sont des variables aléatoires réelles centrées, réduites et non corrélées.

## 3. Co-inertie de deux variables aléatoires hilbertiennes

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Soient  $\mathcal{X}$  et  $\mathcal{Y}$  (supposées centrées) deux variables aléatoires d'ordre 2 sur  $(\Omega, \mathcal{A}, P)$ , à valeurs dans les espaces de Hilbert réels séparables  $H_1$  et  $H_2$ , de produits scalaires  $(\cdot|\cdot)_1$  et  $(\cdot|\cdot)_2$ . D'après la formule (1), leurs sont associées les applications linéaires continues de Hilbert-Schmidt  $u_{\mathcal{X}}$  et  $u_{\mathcal{Y}}$ .

L'analyse de co-inertie de  $\mathcal{X}$  et  $\mathcal{Y}$  est définie comme la recherche des éléments  $a$  de  $H_1$  et  $b$  de  $H_2$  de norme unité tels que  $u_{\mathcal{X}}(a)$  et  $u_{\mathcal{Y}}(b)$  soient de covariance maximum, *et itérations sous contraintes d'orthonormalité dans  $H_1$  et  $H_2$ .*

Les fonctions orthogonales empiriques  $a$  et  $b$  sont les fonctions propres des opérateurs nucléaires auto-adjoints positifs :

$$u_{\mathcal{X}}^{\bullet} u_{\mathcal{Y}} u_{\mathcal{Y}}^{\bullet} u_{\mathcal{X}} a = \lambda a \quad ||a||_1^2 = 1, \quad (6)$$

$$u_{\mathcal{Y}}^{\bullet} u_{\mathcal{X}} u_{\mathcal{X}}^{\bullet} u_{\mathcal{Y}} b = \lambda b \quad ||b||_2^2 = 1. \quad (7)$$

La covariance maximum est alors :

$$E(u_{\mathcal{X}}(a)u_{\mathcal{Y}}(b)) = \sqrt{\lambda} \quad (8)$$

Cette proposition est une généralisation hilbertienne de l'analyse de co-inertie de CHESSEL et MERCIER (1993). La co-inertie présente l'intérêt par rapport à l'analyse canonique de deux variables aléatoires hilbertiennes de ne pas entraîner de problèmes de compacité des opérateurs considérés. Une alternative à la co-inertie de deux ensembles de courbes, dans l'esprit de l'analyse canonique, a toutefois été récemment proposée par LEURGANS *et al.* (1993). On aurait également pu considérer l'ACP conjointe de deux ensembles de courbes de RICE et SILVERMAN (1991).

#### 4. Espaces de Sobolev

Si nous nous limitons à l'étude de courbes définies sur un certain intervalle  $[a, b]$ , il suffira que  $H$  soit un espace de Sobolev  $H^m([a, b])$  pour couvrir la plupart de nos besoins.

$$H^m([a, b]) = \{f \in C^{m-1}([a, b]), f^{(m)} \in L^2([a, b])\} \quad (9)$$

$C^{m-1}([a, b])$  est l'espace des fonctions dont la dérivée  $(m-1)^{\text{ième}}$  est continue sur  $[a, b]$ .  $f^{(m)}$  est la dérivée  $m^{\text{ième}}$  de  $f$ .

Les espaces de Sobolev sont des espaces de Hilbert lorsqu'ils sont munis de produits scalaires du type :

$$(f|g)_{H^m} = \sum_{j=0}^m \alpha_j \int_a^b f^{(j)}(t)g^{(j)}(t)dt \quad (10)$$

avec  $\alpha_0 = 1$  et  $0 \leq \alpha_j \leq 1$ ,  $1 \leq j \leq m$ .

Les analyses abstraites des sections 2 et 3 ne sont pas, sauf exception théorique (par exemple ACP d'un mouvement brownien), réalisables. En pratique, il faut chercher une approximation à la fois dans l'espace des variables aléatoires réelles de carré intégrable puisque seul un nombre fini de courbes a été observé, et dans l'espace de Hilbert  $H^m([a, b])$  puisqu'on ne connaît que quelques points de ces courbes. Afin de les reconstituer, nous allons employer des fonctions splines et plus particulièrement une base de B-splines, comme BOUHADDOU *et al.* (1987). Les bases de B-splines

sont des outils (généralement) différents des bases splines tirées de noyaux auto-reproduisants, utilisées par BESSE et RAMSAY (1986).

### 5. Espace spline et fonctions B-splines\*

Les fonctions splines polynomiales sont des fonctions constituées de morceaux de polynômes se raccordant plus ou moins. On appelle *espace des splines polynomiales* sur l'intervalle  $[a, b]$  d'ordre  $m$  en les nœuds simples  $a = x_0 < x_1 < \dots < x_k < x_{k+1} = b$ , l'espace :

$$\mathbf{S}^m(\Delta) = \{s(t), t \in [a, b]; \exists s_0(t), \dots, s_i(t), \dots, s_k(t) \in \mathbf{P}^{(m)}([a, b])\}$$

tels que  $\forall t \in [x_i, x_{i+1}], s(t) = s_i(t)$  et  $\forall j = 0, 1, \dots, m-2$   
et  $i = 1, \dots, k, s_{i-1}^{(j)}(x_i) = s_i^{(j)}(x_i)$ .

$\mathbf{P}^{(m)}([a, b])$  est l'espace des polynômes d'ordre  $m$  définis sur  $[a, b]$ .

$\mathbf{S}^m(\Delta)$  est un sous-espace hilbertien de  $\mathbf{H}^{m-1}([a, b])$  de dimension  $r = k + m$ . Une base particulièrement intéressante de cet espace est celle des fonctions B-splines polynomiales normalisées  $(B_j^m(t))_{j=1, \dots, r}$  dont il existe plusieurs définitions, par exemple à partir de la formule de DE BOOR-COX-MANSFIELD :

Soit la *partition associée* (automatiquement) à  $\mathbf{S}^m(\Delta) : \tilde{\Delta} = (\delta_i)_{i=1, \dots, m+r}$ , avec  $\delta_1 = \delta_2 = \dots = \delta_m = a$ ,  $\delta_{i+m} = x_i$  pour  $i = 1$  à  $k$  et  $\delta_{r+1} = \delta_{r+2} = \dots = \delta_{r+m} = b$ .

Il s'agit d'une récurrence sur cette partition :

$$B_j^1(t) = \begin{cases} 1 & \text{si } \delta_j \leq t \leq \delta_{j+1} \\ 0 & \text{sinon} \end{cases}.$$

Pour  $s = 2, \dots, m$  :

$$B_j^s(t) = \frac{t - \delta_j}{\delta_{j+s-1} - \delta_j} B_j^{s-1}(t) + \frac{\delta_{j+s} - t}{\delta_{j+s} - \delta_{j+1}} B_{j+1}^{s-1}(t). \quad (11)$$

Les B-splines possèdent un support restreint, sont positives et forment une partition de l'unité.

Leur dérivée  $D(B_j^m(t))$  ( $D$  est l'opérateur de dérivation) s'exprime en fonction des B-splines d'ordre inférieur :

$$D(B_j^m(t)) = (m-1) \left( \frac{B_j^{m-1}(t)}{\delta_{j+m-1} - \delta_j} - \frac{B_{j+1}^{m-1}(t)}{\delta_{j+m} - \delta_{j+1}} \right), \quad (12)$$

\* Les éléments de cette section sont exposés en détail et démontrés dans SCHUMAKER (1981).

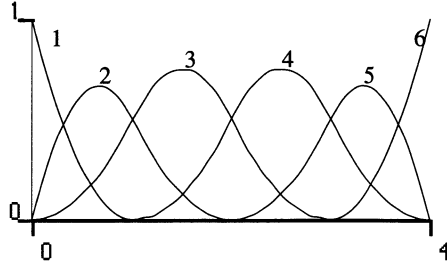


FIGURE 1

Base de 6 B-splines d'ordre 3 définie sur  $[0,4]$  avec trois nœuds intérieurs (1,2,3).

sachant que pour les deux B-splines extrêmes, l'un des termes disparaît. Leur intégrale s'exprime en fonction des B-splines d'ordre supérieur :

$$\int_a^t B_j^m(u) du = \frac{\delta_{j+m} - \delta_j}{m} \sum_{w=j}^r B_w^{m+1}(t). \quad (13)$$

Le produit scalaire de Lebesgue de deux B-splines

$$g_{ij} = \int_a^b B_i^m(t) B_j^m(t) dt \quad (14)$$

se calcule par une quadrature de Gauss. Ces produits scalaires sont organisés dans une matrice  $\mathbf{G}$ , dite de Gram :  $\mathbf{G} = [g_{ij}]_{i,j=1,\dots,r}$ .

L'interpolation et l'approximation aux moindres carrés par les fonctions B-splines se font au moyen des techniques habituelles pour les modèles linéaires en leurs paramètres, les systèmes linéaires en jeu étant bien conditionnés (à bandes).

## 6. M-splines, I-splines, B-splines, étude de diverses contraintes

Les B-splines sont loin d'être les seules bases de splines utilisées dans la littérature statistique. Si l'on excepte les bases tirées de noyaux reproduisants, qui ressortent d'une théorie différente, et que l'on se cantonne aux splines de régression, il est possible de citer les M-splines (CURRY et SCHOENBERG, 1966), les I-splines (RAMSAY, 1988), les fonctions «puissance tronquée» ou fonctions «+» (DE BOOR, 1978)...

### 6.1. Les M-splines

Elles sont définies à partir des B-splines par :

$$M_j^m(t) = \frac{m}{\delta_{j+m} - \delta_j} B_j^m(t). \quad (15)$$

Elles partagent le même support que les B-splines, sont également positives, mais la propriété de partition de l'unité est remplacée par une autre normalisation :

$\int_a^b M_j^m(t) dt = 1$ . Ces propriétés donnent aux M-splines le statut de densité de probabilité.

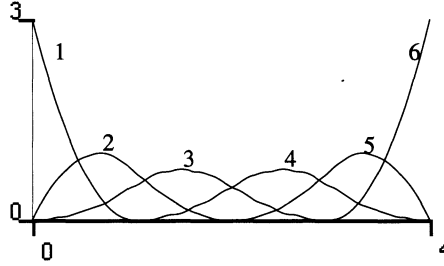


FIGURE 2

Base de 6 M-splines d'ordre 3 définie sur  $[0, 4]$  avec trois nœuds intérieurs (1, 2, 3)

### 6.2. Les I-splines

La définition analytique des I-splines est :

$$I_j^{m+1}(t) = \int_a^t M_j^m(u) du. \quad (16)$$

D'où la monotonie de  $I_j^{m+1}(t)$  sur  $[a, b]$  et les égalités :  $I_j^{m+1}(a) = 0$  et  $I_j^{m+1}(b) = 1$ . Ces fonctions ont par conséquent le statut de fonction de répartition.

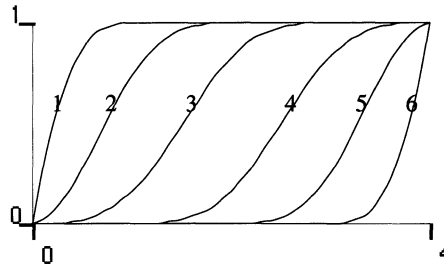


FIGURE 3

Base de 6 I-splines d'ordre 4 définie sur  $[0, 4]$  avec trois nœuds intérieurs (1, 2, 3)

On peut les écrire en fonction des B-splines :

$$I_j^{m+1}(t) = \frac{m}{\delta_{j+m} - \delta_j} \int_a^t B_j^m(u) du = \sum_{w=j}^r B_w^{m+1}(t) \quad (17)$$



### 6.3. Quelques contraintes

Moyennant des contraintes simples sur les coefficients (de positivité ou de somme unitaire), il est possible de donner des conditions suffisantes d'appartenance des combinaisons linéaires des éléments de ces bases à des sous-ensembles de l'espace spline : fonctions positives, fonctions monotones, densités de probabilité, fonctions de répartition...

TABLEAU 1  
Action de différentes contraintes (en colonnes)  
sur trois bases de l'espace spline (en lignes)  
 $f(t)$  représente une combinaison linéaire  
des éléments des bases respectant les contraintes

	sans contrainte	contraintes $a_j \geq 0$	contrainte $\sum_j a_j = 1$
B-spline		$f(t) \geq 0$	
M-spline		$f(t) \geq 0$	$\int_a^b f(t)dt = 1$
I-spline	$f(a) = 0$	$f^{(1)}(t) \geq 0$	$f(b) = 1$

Les contraintes étant linéaires, un algorithme de type gradient projeté convient pour estimer les paramètres aux moindres carrés. Notons également que, du fait du support local de ces fonctions, il est possible d'employer des contraintes de positivité ou de monotonie seulement sur une partie de la courbe (par exemple au début d'une courbe de croissance) en forçant les premiers coefficients I-splines à la positivité. KELLY et RICE (1990) proposent un autre type de contrainte afin de forcer directement les B-splines à la monotonie.

D'autres contraintes telles que symétrie ou périodicité peuvent également être prises en compte grâce aux B-splines.

### 6.4. Retour aux B-splines

La procédure d'estimation des coefficients se déroule donc parfois de préférence dans les bases de M-splines ou I-splines. Des formules permettent de revenir aux B-splines afin de réintégrer les schémas d'analyse proposés dans les deux sections suivantes.

$$\text{Soit } f(t) = \sum_{j=1}^r a_j M_j^m(t) \text{ alors } f(t) = \sum_{j=1}^r \frac{m a_j}{\delta_{j+m} - \delta_j} B_j^m(t). \quad (18)$$

$$\text{Soit } f(t) = \sum_{j=1}^r a_j I_j^{m+1}(t) \text{ alors } f(t) = \sum_{j=0}^r b_j B_j^{m+1}(t) \quad (19)$$

avec  $b_0 = 0$  et  $b_j = b_{j-1} + a_j$  pour  $1 \leq j \leq r$ .

## 7. Approximation B-spline de l'ACP d'une variable aléatoire hilbertienne

On considère à présent une collection de  $n$  courbes splines  $s_i(t)$  de  $\mathbf{S}^m(\Delta)$ , résultant d'une procédure d'interpolation ou de lissage. Leurs coefficients dans la base de B-splines  $(B_j^m(t))_{j=1,\dots,r}$  sont rangés dans une matrice  $\mathbf{S} = [s_{ij}]_{i=1\dots n, j=1\dots r}$ . La probabilité uniforme ( $\mathbf{D} = (1/n)\mathbf{I}_n$ ) est employée sur cet espace fini.  $\mathbf{G}(r \times r)$  est la matrice des produits scalaires de Sobolev entre les éléments de la base de B-splines, calculés en combinant formules (12) et (14).

En reformulant les équations (1) et (2) avec la restriction de  $\mathbf{H}^{m-1}([a, b])$  à  $\mathbf{S}^m(\Delta)$ , isomorphe à  $(\mathbb{R}^r, \mathbf{G})$ , et en employant l'isomorphisme entre  $\mathbf{L}^2(\Omega, \mathcal{A}, P)$  et  $(\mathbb{R}^n, \mathbf{D})$ , les applications linéaires continues  $u$  et  $u^\bullet$  s'expriment comme :

$$\forall \mathbf{h} \in \mathbb{R}^r, \quad u(\mathbf{h}) = \mathbf{S}\mathbf{G}\mathbf{h}. \quad (20)$$

$$\forall \mathbf{z} \in \mathbb{R}^n, \quad u^\bullet(\mathbf{z}) = \mathbf{S}'\mathbf{D}\mathbf{z}. \quad (21)$$

Les fonctions orthogonales empiriques s'obtiennent (dans la base de B-splines) par la diagonalisation de la matrice  $\mathbf{G}$ -symétrique positive :

$$\mathbf{S}'\mathbf{D}\mathbf{S}\mathbf{G}\mathbf{a} = \lambda \mathbf{a} \quad \text{avec} \quad \mathbf{a}'\mathbf{G}\mathbf{a} = 1. \quad (22)$$

Soit  $c$ , composante principale,  $\sqrt{\lambda c} = \mathbf{S}\mathbf{G}\mathbf{a}$  est de  $\mathbf{D}$ -norme maximum.

On peut étudier la convergence de cette analyse vers l'ACP de la variable aléatoire hilbertienne correspondante.

## 8. Approximation B-spline de la co-inertie de deux variables aléatoires hilbertiennes

On considère deux ensembles appariés de  $n$  courbes splines  $h_i(t)$  de  $\mathbf{S}^m(\Delta_1)$  et  $g_i(t)$  de  $\mathbf{S}^l(\Delta_2)$ . Leurs coefficients dans les bases de B-splines  $(B_j^m(t))_{j=1,\dots,r_1}$  et  $(B_k^l(t))_{k=1,\dots,r_2}$  sont rangés dans des matrices  $\mathbf{S}_1 = [s_{1ij}]_{i=1\dots n, j=1\dots r_1}$  et  $\mathbf{S}_2 = [s_{2ik}]_{i=1\dots n, k=1\dots r_2}$ . Les matrices de Gram correspondantes sont  $\mathbf{G}_1 = [g_{1jj'}]_{j,j'=1,\dots,r_1}$  et  $\mathbf{G}_2 = [g_{2kk'}]_{k,k'=1,\dots,r_2}$ . L'espace probabilisé conserve la probabilité uniforme  $\mathbf{D}$ .

En réécrivant les équations de la section 3 dans  $\mathbf{S}^m(\Delta_1) = H_1$ ,  $\mathbf{S}^l(\Delta_2) = H_2$  (dans leurs bases B-splines respectives) et  $(\mathbb{R}^n, \mathbf{D}) \approx \mathbf{L}^2(\Omega, \mathcal{A}, P)$ , on obtient :

$$\forall \mathbf{h} \in \mathbb{R}^{r_1}, \quad u_x(\mathbf{h}) = \mathbf{S}_1\mathbf{G}_1\mathbf{h}, \quad (23)$$

$$\forall \mathbf{g} \in \mathbb{R}^{r_2}, \quad u_y(\mathbf{g}) = \mathbf{S}_2\mathbf{G}_2\mathbf{g}. \quad (24)$$

On cherche à maximiser :

$$\mathbf{g}'\mathbf{G}_2\mathbf{S}_2'\mathbf{D}\mathbf{S}_1\mathbf{G}_1\mathbf{h} \quad \text{avec} \quad \mathbf{h}'\mathbf{G}_1\mathbf{h} = 1 \quad \text{et} \quad \mathbf{g}'\mathbf{G}_2\mathbf{g} = 1. \quad (25)$$

Ces vecteurs  $h$  et  $g$  découlent de la décomposition du triplet statistique (ESCOUFIER, 1987) ( $S_2DS_1, G_1, G_2$ ). On peut étudier la convergence de cette analyse vers l'analyse abstraite de la section 3.

## 9. Application à l'étude de la zone forestière d'Amance

Pour mieux appréhender la dynamique des peuplements forestiers, il est essentiel d'étudier l'impact de la structure du peuplement à la fois sur la croissance en circonférence et en hauteur des arbres. En ce qui concerne les peuplements plantés (équiennes\*\* et souvent monospécifiques), une des variables explicatives de cette structure est le nombre de tiges à l'hectare : la densité. L'impact de la densité est différent pour la hauteur et pour la circonférence : il est très largement admis que la croissance en circonférence diminue avec la densité (ASSMANN, 1970; DREYFUS, 1990) tandis que la croissance en hauteur y paraît beaucoup moins sensible, avec des conclusions parfois contradictoires. Les fortes densités peuvent avoir un effet positif, poussant les arbres à favoriser leur croissance en hauteur pour rester dans l'étage dominant. ASSMANN (1970) précise que l'Épicéa est une essence particulièrement sensible à de fortes densités pour sa croissance en hauteur. Mais une forte densité a aussi un effet négatif sur la croissance, en limitant les ressources disponibles par arbre (DREYFUS, 1990). Les études à ce sujet restent rares, les mesures non destructives de hauteur étant plus délicates à effectuer que celles de la circonférence. Il semble donc intéressant d'étudier conjointement l'évolution de ces deux croissances.

### 9.1. Le dispositif expérimental

Les mesures ont été faites sur des arbres abattus dans un clinal d'Épicéas communs (provenance polonaise Istebna) de la forêt d'Amance, près de Nancy. Ce clinal avait été mis en place en 1970 par l'INRA, avec un gradient continu de densité du nord au sud, sur 60 mètres (les arbres sont en croissance libre au nord, et en forte densité au sud; un descriptif complet du dispositif peut être trouvé dans DREYFUS, 1990). Mise à part cette contrainte de densité, les arbres sont répartis aléatoirement. Les arbres placés à l'ouest sont proches de la lisière et sujets à des effets de bord.

En début d'année 1993, une très forte éclaircie systématique a été faite, touchant des arbres de tous les statuts sociaux dans les différentes densités. Elle a permis de réunir des données appariées sur les hauteurs (mesures des croissances annuelles grâce aux cicatrices le long du tronc, de 1974 à 1990) et les circonférences (mesures des cernes annuels d'une rondelle, de 1974 à 1992), pour plus d'une centaine d'arbres d'une vingtaine d'années.

---

\*\* terme forestier signifiant que les arbres sont tous du même âge.

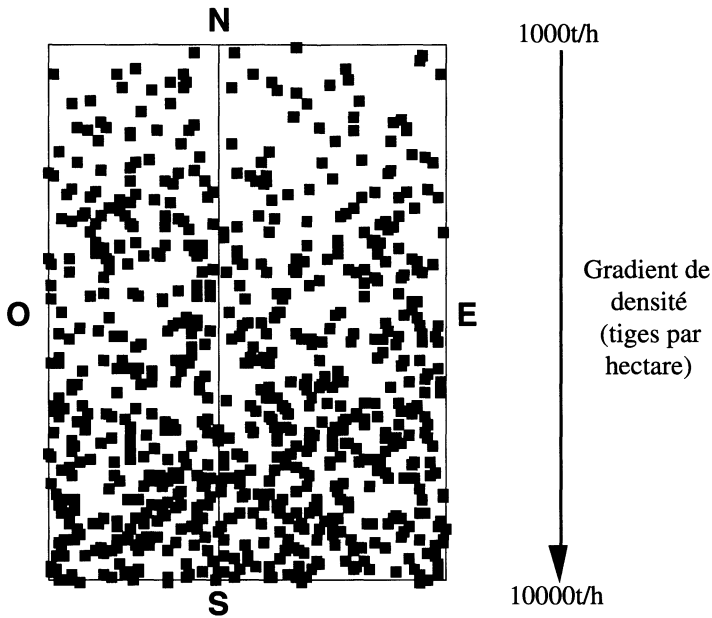


FIGURE 4

*Dispositif d'Amance, chaque carré représente un arbre planté*

### 9.2. Courbes de croissance en hauteur et en circonférence

Des I-Splines monotones d'ordre 4 ont été employées pour modéliser les croissances cumulées (EL FAOUZI et ESCOUFIER, 1991) en hauteur et en circonférence, et rendent compte des tendances générales.

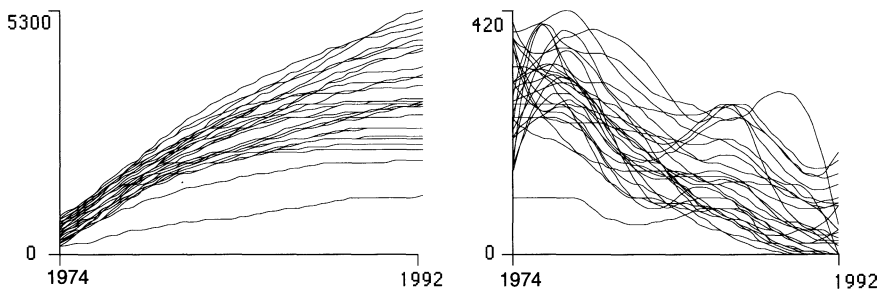


FIGURE 5

*A gauche, modélisation I-spline de la croissance en circonférence d'une trentaine d'Épicéas sélectionnés aléatoirement (sur 141); à droite, dérivées de ces fonctions*

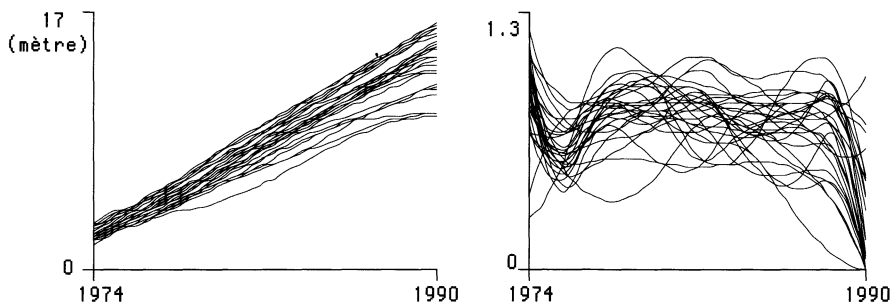


FIGURE 6

*A gauche, modélisation I-spline de la croissance en hauteur  
d'une trentaine d'Épicéas sélectionnés aléatoirement;  
à droite, les dérivées de ces fonctions*

La croissance en circonférence a une forme parabolique, ce qui correspond à une augmentation linéaire de la surface terrière (surface de la section de tronc à 1m30).

La croissance en hauteur est dans sa phase linéaire, les arbres étant relativement jeunes, ils n'ont pas encore atteint le palier classique de croissance. Les accroissements montrent que la hauteur a été plus durement touchée que la circonférence par la sécheresse de 1976.

### 9.3. Analyse de co-inertie

Il est courant en foresterie d'étudier les taux de croissance de préférence à la croissance elle-même (HOULLIER *et al.*, 1991). L'analyse de co-inertie sera effectuée sur les taux de croissance. Pour ce faire, les I-splines estimées sur les données cumulées sont d'abord transformées en B-splines (formule(19)) afin de se replacer dans les schémas d'analyse définis. On obtient alors les taux de croissance en dérivant simplement les fonctions B-splines de croissance (formule (12)).

L'objectif de l'analyse est de générer deux codes a-temporels de covariance maximum, résumant les taux de croissance en hauteur et en circonférence. Maximiser la covariance est un compromis entre la maximisation des variances de chacun des codes (objectif de l'ACP pour la description de structures) et la maximisation de la corrélation (objectif de l'analyse canonique pour la description de co-structure).

Pour une comparaison graphique directe, les codes appariés sont normalisés et sont appelés «pseudo-codes canoniques» pour rappeler la part d'analyse canonique qu'ils contiennent. Comme le suggèrent JONES et RICE (1992) pour l'ACP, on peut sélectionner les valeurs extrêmes de ces codes et représenter les courbes correspondantes (et leurs dérivées) pour interpréter les pseudo-codes. On exprime ainsi clairement un aspect du fouillis des figures 5 et 6. La cartographie de ces pseudo-codes canoniques permet ensuite l'interprétation spatiale des variations.

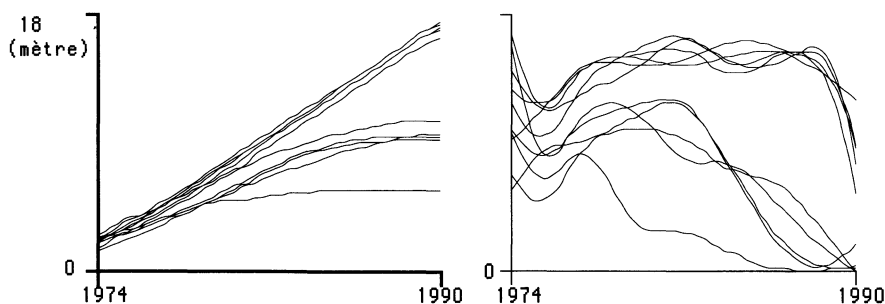


FIGURE 7

*A gauche, représentation des cinq courbes de hauteur dont le premier pseudo-code canonique présente les valeurs minimales et des cinq courbes avec les valeurs maximales; à droite, les dérivées de ces fonctions*

#### 9.4. Premier code canonique

Le premier code extrait plus de 95% de la variabilité et traduit un effet taille tout au long de la croissance (voir figure 7, la représentation des courbes pour la circonférence serait identique) : les courbes ne se coupent pas et le nuage de courbes se disperse. Ce résultat est classique pour la circonférence, moins en ce qui concerne la hauteur.

La répartition spatiale des codes (Figure 8) montre une très forte structuration le long du gradient de densité pour la hauteur et pour la circonférence; les arbres ayant la plus forte croissance sont en faible densité. Les deux structures présentent les mêmes regroupements qui doivent correspondre à des zones de densité locale homogène. Ceci est particulièrement net pour la partie non soumise aux effets de bord de la lisière ouest.

Cette forte structure commune est le reflet de la répartition sociale des arbres : des grands arbres dominants ont un avantage pour la photosynthèse et donc la croissance. La pression sociale augmentant avec la densité, il est logique de trouver les plus faibles croissances en hauteur et en circonférence du côté de la forte densité. En plus de cet effet du statut social, la distribution des fortes croissances en hauteur montre que la densité a un effet négatif sur la croissance en hauteur des arbres dominants. Ceci est très net au delà d'une densité de 4000 t/ha, mais le passage de 1000 t/ha à 4000 t/ha sur 20 mètres, ne permet pas d'analyser plus finement la sensibilité de la hauteur à la densité.

#### 9.5. Second code canonique

Le deuxième code extrait 3% de la variabilité. Il permet de séparer deux tendances : une croissance régulière pour les codes les plus faibles ou avec un palier pour les codes les plus forts (voir figure 9, là encore seule la hauteur est représentée).

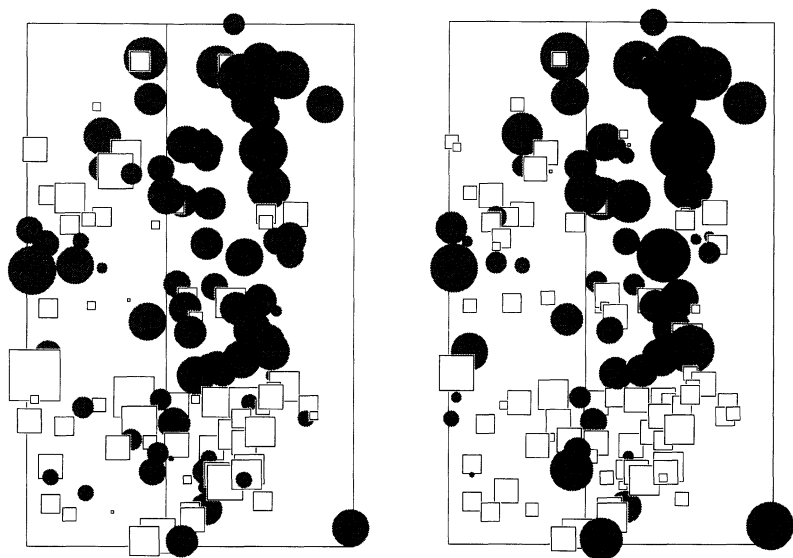


FIGURE 8

*Cartographie du premier pseudo-code canonique, à gauche la hauteur et à droite la circonférence.*

*Les ronds (resp. carrés) correspondent à des valeurs positives (resp. négatives) de surface proportionnelle*

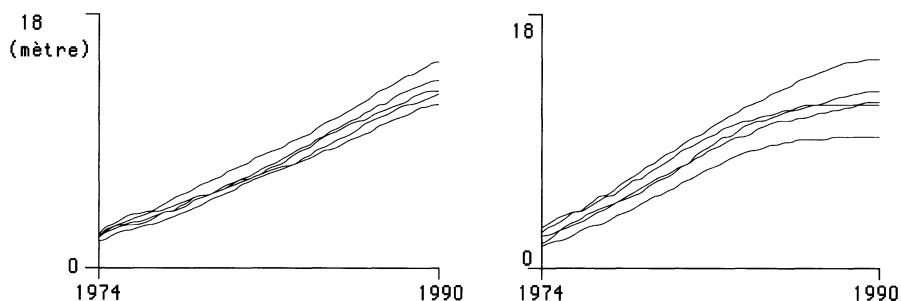


FIGURE 9

*A gauche, représentation des cinq courbes de hauteurs dont le second pseudo-code canonique présente les valeurs minimales et à droite les cinq courbes avec les valeurs maximales*

A part deux groupes homogènes (au nord-ouest et sud-est), les deux répartitions spatiales (Figure 10) ne présentent pas les mêmes similitudes que pour le premier code. Pour la croissance en hauteur, il se pourrait que le long du bord ouest on retrouve un effet lisière (meilleur éclaircissement). Pour la circonférence, la structure le long du gradient de densité est à nouveau très forte, les arbres en faible densité ayant plutôt une croissance en palier.

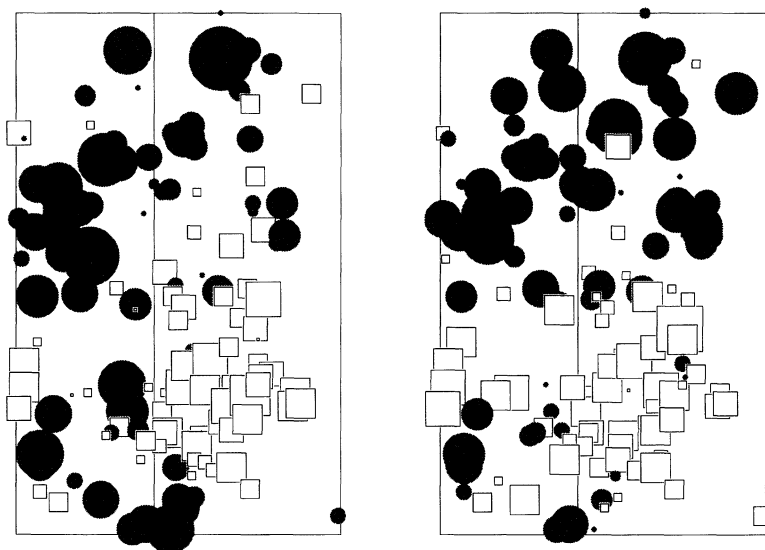


FIGURE 10

*Cartographie du second pseudo-code canonique,  
à gauche la hauteur et à droite la circonférence*

Il est possible d'interpréter ces résultats en termes de fermeture du couvert forestier et d'installation de relations de compétition. En effet les arbres en faible densité sont quasiment en croissance libre les premières années, le couvert forestier se refermant assez tardivement, puis leur croissance ralentit quand les houppiers entrent en concurrence. Au contraire le couvert se referme très rapidement en forte densité et la croissance des arbres est dès le départ ralentie et régulière.

## 10. Conclusion

Il est difficile de poursuivre plus loin ces interprétations car il faudrait plus de données sur une plus longue période de croissance, en particulier pour voir clairement le ralentissement de la croissance en hauteur. Cette première étude montre l'intérêt dans le domaine forestier d'un outil permettant d'analyser simultanément deux aspects de la croissance des arbres : celle en hauteur et celle en circonférence. Remarquons à nouveau qu'il respecte certaines contraintes naturelles de ces croissances, comme la monotonie. Par rapport à la proposition d'HOULLIER (1987), cette approche semi-paramétrique, outre qu'elle permet un résultat explicite de convergence, ne nécessite pas l'introduction *a priori* d'un modèle de croissance et est en ce sens beaucoup plus « libre ». Cette liberté provient du fait que *la courbe sort des données*. Il est donc impératif que les mesures soient réparties sur l'ensemble de l'intervalle considéré. Dans le cas où, pour certaines courbes, les données sont localisées sur une portion de l'intervalle, *la courbe doit sortir d'un modèle*, si toutefois on peut l'exhiber. L'ACP sur paramètres est alors une solution très intéressante. De plus, elle



permet une interprétation directe des résultats en terme de paramètres (peu nombreux) possédant une signification biologique. On peut d'ailleurs concevoir facilement une généralisation de l'ACP sur paramètres d'HOULLIER à une co-inertie de deux modèles différents appariés dans l'esprit de la section 8.

### Remerciements

Nous tenons à remercier F. COLIN, Ph. DREYFUS et F. NINGRE de l'INRA de nous avoir permis l'acquisition des données dans le dispositif clinal, ainsi que F. HOULLIER pour ses remarques constructives sur le manuscrit de cet article.

### Références

- [1] ASSMAN E. (1970) *The principles of forest yield study. Studies in the organic production, structure, increment and yield of forest stands*. Pergamon Press : Oxford.
- [2] BESSE Ph. et RAMSAY J.O. (1986) Principal components analysis of sampled functions. *Psychometrika*, 51, 285-311.
- [3] BOUHADDOU O., OBLED C.H. et DINH T.P. (1987) Principal components analysis and interpolation of stochastic processes : method and simulation. *Journal of applied statistics*, 14, 251-267.
- [4] CURRY H.B. et SCHOENBERG I.J. (1966) On Polya frequency functions and their limits. *Journal d'Analyse Mathématique*, 17, 71-107.
- [5] CHESSEL D. et MERCIER P. (1993) Couplage de triplets statistiques et liaisons espèces-environnement. In *Biométrie et environnement*, J.D. Lebreton et B. Asselin (Eds), Société Française de Biométrie, Masson : Paris, 15-45.
- [6] DAUXOIS J. et POUSSE A. (1976) Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique. *Thèse de troisième cycle*, Université Paul Sabatier, Toulouse.
- [7] DE BOOR C. (1978) *A practical guide to splines*. Springer-Verlag : New-York.
- [8] DE BOOR C. et DANIEL J.W. (1974) Splines with non negative B-spline coefficients. *Mathematics of computation*, 28, 565-568.
- [9] DREYFUS P. (1990) Interaction génotype-densité et compétition dans un dispositif clinal d'épicéas communs. *Annales des Sciences Forestières*, 47, 1-16.
- [10] EL FAOUZI N. et ESCOUFIER Y. (1991) Modélisation I-spline et comparaison de courbes de croissance. *Revue de Statistique Appliquée*, 39, 51-64.
- [11] ESCOUFIER Y. (1987) The duality diagram : a means for better practical applications. In *Developments in numerical ecology*. P. Legendre et L. Legendre (eds). NATO Advanced Study Institute Series G (Ecological Sciences). Springer-Verlag : Berlin, 139-156.

- [12] GASSER T., MÜLLER H.G., KÖHLER W., MOLINARI L. et PRADER A. (1984) Nonparametric regression analysis of growth curves. *The Annals of Statistics*, 12, 210-229.
- [13] HOULLIER F. (1987) Comparaison de courbes et de modèles de croissance. *Statistique et Analyse des Données*, 12, 17-36.
- [14] HOULLIER F., BOUCHON J. et BIROT Y. (1991) Modélisation de la dynamique des peuplements forestiers : état et perspectives. *Revue Forestière Française*, 2, 87- 108.
- [15] JONES M.C. et RICE J.A. (1992) Displaying the important features of large collections of similar curves. *The American Statistician*, 46, 140-145.
- [16] KELLY C. et RICE J.A. (1990) Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, 46, 1071-1085.
- [17] LEURGANS S.E., MOYEED R.A. et SILVERMAN B.W. (1993) Canonical correlation analysis when the data are curves. *Journal of Royal Statistical Society B*, 55, 725-740.
- [18] LOÈVE M. (1963) *Probability theory*. Van Nostrand : Princeton.
- [19] RAMSAY J.O. (1988) Monotone regression splines in action. *Statistical Science*, 3, 425-461 (avec discussion).
- [20] RICE J.A. et SILVERMAN B.W. (1991) Estimating the mean and covariance structure non-parametrically when the data are curves. *Journal of Royal Statistical Society B*, 53, 233-243.
- [21] SCHUMAKER L. (1981) *Spline functions : Basic theory*. Wiley : New York.
- [22] TUCKER L. R. (1958) An inter-battery method of factor analysis. *Psychometrika*, 23, 111-136.