STATISTIQUE ET ANALYSE DES DONNÉES

PHILIPPE BESSE
CHRISTINE THOMAS-AGNAN

Le lissage par fonctions splines en statistique, revue bibliographique

Statistique et analyse des données, tome 14, n° 1 (1989), p. 55-84.
<http://www.numdam.org/item?id=SAD_1989__14_1_55_0>

Statistique et Analyse des Données
1989 - Vol. 14 n° 1 - p. 55 - 84

LE LISSAGE PAR FONCTIONS SPLINES EN STATISTIQUE

REVUE BIBLIOGRAPHIQUE

Philippe **BESSE**, Christine **THOMAS-AGNAN**

Laboratoire de Statistique et Probabilités
U.A. C.N.R.S. 745
Université Paul Sabatier
31 062 TOULOUSE Cedex.

Résumé : *Cet article bibliographique a pour objet de présenter l'utilisation des fonctions splines dans divers domaines de la Statistique, plus particulièrement la régression non-paramétrique et la prédiction de processus. Il aborde le problème du choix de la valeur du paramètre de lissage. Pour permettre au lecteur de manipuler cet outil, cette présentation est précédée d'une introduction à la théorie des fonctions splines avec des références à la littérature d'Analyse Numérique dans le cadre de laquelle cette théorie s'est d'abord développée.*

Mots-clés : fonctions splines, lissage, régression non-paramétrique, validation croisée.

Abstract : *In this paper, we review the statistical applications of spline functions theory, with emphasis on nonparametric regression and prediction of processes. We consider the different techniques for choosing a smoothing parameter from the data. To help the reader in using splines as a tool, we give a short background on the theory with references to the numerical analysis literature in which it originated.*

Keywords : spline functions, smoothing, nonparametric regression, cross-validation.

Classification STMA : 07 160, 07 120.

Classification AMS : 62 J 99, 65 D 07.

Manuscrit reçu le 28. 11. 88, révisé le 1. 3. 89

1 - INTRODUCTION

Le mot anglais "spline" désigne une latte flexible utilisée par les dessinateurs pour matérialiser des lignes à courbure variable et passant par des points fixés a priori ou à "proximité" de ceux-ci. Le tracé ainsi réalisé minimise l'énergie de déformation de la latte. Par analogie, ce mot désigne également des familles de fonctions d'interpolation ou de lissage présentant des propriétés "optimales" de régularité. L'idée originale est attribuée à Whittaker(1923), puis reformulée par Schoenberg(1964), Atteia(1965) et Reinsch(1967). Elle connaît ses premières applications en Statistique avec Kimeldorf et Wahba(1970).

Soit f une fonction réelle définie sur un intervalle compact qui, pour simplifier, est [0 , 1]. Soient $\{t_i ; i = 1,...,n\}$ n points distincts de [0 , 1] et $y_i = f(t_i) + \varepsilon_i$ des observations entachées d'erreur de f. Lorsque f est supposée appartenir à un espace défini par un nombre fini de paramètres, son estimation relève des techniques de régression classiques. Pour s'affranchir de cette hypothèse, on peut simplement supposer une certaine régularité de f c'est-à-dire l'existence de ses dérivées jusqu'à un certain ordre ou encore, l'appartenance à un espace de Sobolev:

$$f \in W^{(m)} = \left\{ f \in C^{(m-1)}[0,1] \mid f^{(m)} \in L^2[0,1] \right\}.$$

De nombreuses présentations sont développées dans la littérature mais la définition conduisant aux splines "naturelles" modélisant l'outil du dessinateur (pour m = 2) consiste à estimer f par la solution du problème de minimisation :

$$\underset{h \in W^{(m)}}{\text{Min}} \frac{1}{n} \sum_{i=1}^{n} [y_i - h(t_i)]^2 + \lambda \int_0^1 [h^{(m)}(t)]^2 \, dt \qquad (1.1)$$

La constante de lissage ($\lambda > 0$) détermine le compromis entre la fidélité aux données et la régularité de la solution. On montre alors que, dans ce cas, la solution est constituée de morceaux de polynômes de degré (2m-1) entre les points t_i (appelés nœuds) qui se raccordent de façon à satisfaire les hypothèses de régularité (la dérivée d'ordre (2m-1) est une fonction étagée).

Les familles de splines constituent un ensemble d'outils d'approximation, de lissage et d'interpolation largement utilisés dans des domaines d'application des mathématiques très variés (problèmes de régularisation, d'équations intégrales, de restauration d'images, ...) aussi la bibliographie sur le sujet est considérable. Une liste

complète des articles parus jusqu'en 1973 sur les splines a été publiée par Van Rooij et Schurer(1974) et Wegman et Wright(1983) en résument les applications statistiques. Comme il serait vain de prétendre à l'exhaustivité sur un tel thème dans l'espace d'une publication, le champ d'étude de cet article est restreint aux références directement liées à des applications statistiques hormis quelques outils de base très généraux présentés dans le paragraphe suivant. D'autre part, pour alléger la rédaction, certaines hypothèses très techniques ne sont pas toujours précisées ; le lecteur est invité à se reporter aux références originales.

Le paragraphe trois applique ces outils à la régression non-paramétrique, le quatre aux problèmes de filtrage et de prédiction. Le cinq aborde les problèmes liés au choix du paramètre de lissage et le six résume les spécificités algorithmiques de ces techniques. Enfin, le paragraphe sept est consacré à d'autres types d'applications dont l'analyse en composantes principales de processus, les utilisations "paramétriques" des splines pour les modèles additifs et la réduction de dimension, l'estimation non-paramétrique de la densité.

2 - LES OUTILS

Schématiquement, on peut dire que deux présentations se sont développées parallèlement de part et d'autre de l'Atlantique :

La première approche considère a priori des fonctions définies par morceaux sur les intervalles $[t_i , t_{i+1}]$ avec des conditions de recollement sur la fonction et ses dérivées aux nœuds t_i ainsi que diverses conditions supplémentaires suffisantes pour assurer l'unicité et conduisant à différents types de splines (cf. De Boor(1978) et Schumaker(1981)).

Le deuxième type de présentation s'écrit de façon plus abstraite dans des espaces de Hilbert (Atteia(1965), Laurent(1972)) comme la résolution d'un problème variationnel. Elle généralise celle vue en introduction (1.1) et exprime toujours un compromis entre la fidélité aux données et la régularité de la solution.

Ces deux approches se rejoignent, par exemple, sur les splines "naturelles" (1.1) mais divergent dans leurs généralisations. Ainsi, dans la première approche plus algorithmique, les conditions de recollement peuvent varier d'un nœud à l'autre, des morceaux de polynômes du second degré peuvent être considérés (splines

paraboliques),... . En revanche, la deuxième présentation plus théorique, peut s'étendre à des pavés de $\mathbb{R}^d$ ("Plaques Minces" sur $\mathbb{R}^2$ Duchon(1976)). Nous détaillons ici essentiellement la deuxième présentation.

2.1 - Notations et Définitions

Soient H_1, H_2 deux espaces de Hilbert sur le corps K (qui peut être $\mathbb{R}$ ou $\mathbb{C}$), munis des produits scalaires $(.,.)_1$ et $(.,.)_2$. On note h_1, h_2, ..., h_n, n éléments linéairement indépendants de H_1 et B l'application qui à f appartenant à H_1 associe le n-uplet $\{(f, h_1)_1, (f, h_2)_1, \ldots, (f, h_n)_1\}$ de K^n, lui-même muni du produit scalaire canonique noté (.,.). Il est possible d'introduire des pondérations sur les nœuds en munissant K^n d'une métrique diagonale particulière. Soit T une application linéaire continue de H_1 vers H_2. Soient $Y = {}^t[y_1, y_2, \ldots, y_n]$ un élément de K^n (les données) et λ un réel positif (le paramètre de lissage). Les données sont des mesures, entachées d'erreur dans le cas du lissage, des fonctionnelles $(f, h_i)_1$, et l'on cherche à approximer f par y_λ ainsi définie :

a - lissage (ou ajustement)

y_λ est solution du problème de minimisation :

$$\text{(2.1)} \qquad \min_{h \in H_1} \|Bh - y\|^2 + \lambda \|Th\|_2^2$$

Si l'on introduit les hypothèses :

A1 : si Bh = 0 et Th = 0, alors h = 0,

A2 : l'image de T est fermée dans H_2,

alors le problème admet une solution unique donnée par l'expression :

$$\text{(2.2)} \qquad y_\lambda = \left(B^*B + \lambda T^*T\right)^{-1} B^*Y$$

où M^* désigne l'adjoint d'un opérateur M.

Dans l'espace $K^n \times H_2$ muni du produit scalaire $(.,.) + \lambda(.,.)_2$, le problème (2.1) se réduit à la projection de l'élément z = (y,0) sur l'image de l'application U qui à un élément f de H_1 associe (Bf , Tf), ou encore à la minimisation de $\|z - Uf\|^2_{K^n \times H_2}$.

b - interpolation

Il s'agit cette fois de minimiser $\|Th\|_2^2$ sous les contraintes $h \in H_1$ et $Bh = Y$. On montre que c'est un cas particulier limite du problème (2.1) lorsque λ tend vers 0. Sous les hypothèses A1 et A2, ce problème admet également une solution unique.

A l'autre extrême, pour λ infini, y_λ est l'élément du noyau de T (nécessairement de dimension finie à cause de A1) qui ajuste Y par la méthode des moindres carrés. Ainsi, le lissage par une spline définit une situation intermédiaire entre l'interpolation et les moindres carrés paramétriques usuels sur l'espace de dimension finie KerT. La constante de lissage λ joue donc un rôle fondamental dans la qualité de l'estimation de f.

c - Cas particulier important

Lorsque les données sont les mesures des valeurs d'une fonction f en des points donnés $t_1, t_2, ..., t_n$ (réels), pour assurer l'existence des éléments $h_1, h_2, ..., h_n$ de H tels que $(h_i, f)_1 = f(t_i)$, il est nécessaire et suffisant de supposer que l'espace de Hilbert H_1 est un espace de fonctions d'une variable réelle du type espace de Hilbert à noyau autoreproduisant, c'est-à-dire un espace dans lequel la fonctionnelle δ_t qui à h de H associe h(t) est continue pour tout réel t . Dans un tel espace, il existe donc des éléments e_t tels que $(e_t , h)_1 = h(t)$ (théorème de Riesz) et l'on appelle, noyau reproduisant R(s,t), la fonction de deux variables réelles définie par :

$$R(s,t) = (e_t , e_s)_1 = e_s(t) = e_t(s). \tag{2.3}$$

Dans ce cas, on désignera par A_λ la matrice de l'endomorphisme de K^n qui au vecteur des données Y associe le vecteur $Y_\lambda = By_\lambda$.

2.2 - Exemples

En prenant des espaces et fonctionnelles "concrets", on obtient les divers types de splines usuelles avec, dans chaque cas, $\| Th \|_2^2$ mesurant la régularité de la fonction h.

a - Les D^m-splines

H est l'espace de Sobolev $W^{(m)}$ des fonctions définies sur l'intervalle $I = [0 , 1]$ dont la dérivée d'ordre m (entier) est une fonction de carré intégrable sur I; T est l'opérateur D^m de H sur $L^2(\mathbb{R})$ qui à h associe sa dérivée d'ordre m (cf. Reinsch(1967), Wahba(1975)).

C'est l'exemple de l'introduction (1.1) et le cas le plus usité est celui des splines cubiques (m = 2). Elles appartiennent à la famille des splines polynomiales dans la présentation américaine, c'est-à-dire sont constituées de morceaux de polynômes de degré (2m-1) entre t_1 et t_n et de degré (m-1) en dehors de $[t_1 , t_n]$ permettant de les prolonger à tout $\mathbb{R}$. Ces morceaux sont raccordés aux points t_i de façon à avoir (2m-2) dérivées continues.

Dans le cas où $H = W^{(m)}_{per}$, l'espace des fonctions appartenant à $W^{(m)}$ et périodiques, les solutions deviennent des D^m -splines périodiques (cf Wahba (1975), Wahba et Wold (1975)).

b - Les splines Tchebycheffiennes (ou L-splines)

Soit L un opérateur différentiel linéaire d'ordre m (entier) défini par :

$$L = D \frac{1}{a_1} D \frac{1}{a_2} \dots D \frac{1}{a_m} , \tag{2.4}$$

où les fonctions a_i sont strictement positives et continues jusqu'à l'ordre i {i=1,...,m}, sur l'intervalle I= [0 , 1]. Soit H_L l'espace des fonctions dont les dérivées jusqu'à l'ordre m-1 sont absolument continues, et telles que Lf soit de carré intégrable sur I. L'opérateur L de H_L sur $L^2(\mathbb{R})$ joue alors le rôle de T (Kimeldorf et Wahba(1971)). La forme des fonctions entre les nœuds est donnée par le noyau reproduisant de H_L.

c - Les splines du type Plaques Minces

C'est une extension multidimensionnelle sur $\mathbb{R}^d$ (cf. Duchon(1976), Meinguet(1979)). Le problème s'écrit de manière analogue à (2.1) mais avec

$$\| Th \|_2^2 = \sum_{\Sigma k_i = m} \frac{m!}{k_1! \ldots k_d!} \int_{\mathbb{R}^d} \left(\frac{\partial^m h}{\partial_{x_1}^{k_1} \ldots \partial_{x_d}^{k_d}} \right)^2 dx_1 \ldots dx_p ,$$

l'espace fonctionnel adéquat pour la minimisation étant du type Beppo-Levi (plutôt que Sobolev cf. Deny et Lions(1954)). Ce n'est pas la seule extension multidimensionnelle. Lorsque d=1, la spline coïncide avec la D^m-spline sur tout intervalle contenant les noeuds (et en dehors avec des polynômes de degré (m-1)).

d - Les α-splines

La régularité d'une fonction h est liée au comportement asymptotique des coefficients de Fourier h_n de h si h est périodique, ou , sinon, de la transformée de Fourier $\hat{h}$ de h dans certains cadres fonctionnels. On peut donc construire une mesure de régularité à l'aide de ces quantités.

i) cas périodique :

Soit $\alpha = (\alpha_k)$ une suite de réels positifs et soit H_α l' ensemble des fonctions h de carré intégrable sur I = [0 , 1] telles que :

$$\sum_k |\alpha_k h_k|^2 < \infty \quad .$$

Sous certaines hypothèses concernant la suite α_k, l'espace H_α est un espace de Hilbert à noyau autoreproduisant (cf. Thomas-Agnan(1988)). L'opérateur T de H_α sur l^2 associe à une fonction h la suite $(\alpha_k h_k)$. Cette définition inclut les D^m-splines périodiques comme cas particulier lorsque α_k est proportionnel à k^m.

ii) cas non périodique :

Bien que les espaces utilisés ici soient des espaces de fonctions, il faut avoir recours aux distributions tempérées de Schwartz (espace $\mathcal{S}'(\mathbb{R})$), dans lequel existe la

transformation de Fourier, pour définir aisément les α-splines. Pour une fonction réelle positive $\alpha(.)$ satisfaisant certaines hypothèses, l'ensemble H_α des éléments h de $\mathcal{S}'(\mathbb{R})$ tels que $\alpha(.)\mathcal{F}(.)$ est un élément de $L^2(\mathbb{R})$, est un espace de Hilbert à noyau autoreproduisant (cf. Thomas-Agnan(1988)). T est ici l'opérateur de H_α sur $L^2(\mathbb{R})$ qui à h associe $\alpha(.)\mathcal{F}(.)$. On obtient comme cas particulier les splines Tchebycheffiennes lorsque l'opérateur différentiel L est à coefficients constants et que son polynôme caractéristique n'a pas de racine réelle.

2.3 - Autres définitions

a - B-splines

Pour un espace de splines donné, on donne le nom de B-splines à des splines constituant une base de cet espace et dont chaque élément admet un support restreint au voisinage d'un nœud. Les B-splines existent pour divers types de splines, les plus fréquemment rencontrées étant les B-splines polynômiales, et avec diverses normalisations (cf. Schumaker(1981) de Boor(1978)).

b- Splines Complètes

Lorsque les données comportent des mesures sur les valeurs des dérivées de la fonction à approcher aux extrémités de l'intervalle, celles-ci sont intégrées au terme des moindres carrés dans l'expression à minimiser. Les splines complètes possèdent de meilleures propriétés d'approximation (cf. Schumaker et Utreras(1988)).

c - Splines robustes

Huber(1979) propose de remplacer le terme des moindres carrés de la définition (1.1) :

$$\sum_{i=1}^{n}[y_i - h(t_i)]^2$$

par une expression moins sensible à la présence de données aberrantes et faisant intervenir une fonction convexe ρ :

$$\sum_{i=1}^{n}\rho(y_i - h(t_i)).$$

Comme dans d'autres domaines de la Statistique robuste, ρ peut être défini par:

$$\rho(x) = \begin{cases} x^2 & \text{si } |x| \le c \\ c(2|x| - c) & \text{si } |x| \ge c \end{cases}$$

Utreras(1981a) puis Cox(1983) étudient les propriétés asymptotiques de ce type de spline.

d - Déconvolution et autres problèmes inverses

Lorsque l'on n'est plus dans le cas particulier 2.1.c, les mesures ne sont pas des valeurs aux points t_i de la fonction à estimer, mais des fonctionnelles h_i plus complexes qui peuvent prendre diverses formes selon la nature du problème. C'est le cas par exemple dans la résolution des systèmes différentiels linéaires (cf. Nashed et Wahba(1974), Wahba(1977)). Un autre cas fréquent est celui de la déconvolution lorsque l'on mesure une moyenne pondérée des valeurs de la fonction au voisinage des points t_i (cf. Rice et Rosenblatt(1983)).

e - Inf-convolution splines ou "partial splines"

La fonction à approcher n'est plus entièrement lisse mais peut présenter des singularités de type connu (paramétrisable) a priori, par exemple, une discontinuité en un point connu. On décompose alors f(t) en :

$$f(t) = x(t) + \sum_k \lambda_k p_k(t), \tag{2.5}$$

où x(.) appartient à H_1, espace de fonctions lisses, et $p_k(.)$ paramétrise les singularités. Le problème se ramène à la minimisation de :

$$\|Bh - y\|^2 + \lambda \| Tx \|_2^2, \tag{2.6}$$

où $h = x + \sum_k \lambda_k \ p_k$, $x \in H_1$ et $\lambda_k \in \mathbb{R}$ (cf. Girard (1988), Cox et al.(1988)).

Cette approche est utilisée dans les modèles semi-paramétriques en régression (Heckman(1986), Wahba(1986), Girard(1988)).

f - Splines définies sur un ensemble convexe

Dans certaines situations, le vecteur de données Y est remplacé par une information plus vague du type intervalle. On formalise ceci en envisageant un convexe fermé non vide Γ. Pour le lissage, le problème (2.1) devient la minimisation de :

$$\| Bh - \rho \|_2^2 + \lambda \| Th \|_2^2 ,$$

sous les contraintes $h \in H_1$ et $\rho \in \Gamma$ et, pour l'interpolation, la minimisation de $\| Th \|_2^2$ sous les contraintes $h \in H_1$ et $Bh \in \Gamma$. Sous certaines hypothèses, ces problèmes admettent une solution unique (cf. Atteia(1968)). Ces types de splines sont utilisés par exemple par Wright et Wegman(1983) qui ajoutent des contraintes de monotonie (spline isotone) et Biritxinaga(1987).

2.4 - La matrice Ω et ses valeurs propres

On peut définir une matrice Ω par :

$$(2.7) \qquad {}^tY \Omega Y = \min_{h \in H_1 / (h_i;h)=y_i} \| Th \|_2^2 .$$

Il est aisé de vérifier que $A_\lambda = (Id + \lambda \Omega)^{-1}$. Les valeurs propres μ_k et vecteurs propres v_k de Ω jouent un rôle très important dans beaucoup de problèmes asymptotiques relatifs aux fonctions splines. Ces éléments ne sont pas calculables en général mais peuvent être approchés dans certains cas (cf. Utreras(1983), Schumaker et Utreras(1988)). Pour les D^m-splines, Demmler et Reinsch(1975) montrent que Ω est une matrice d'oscillation, ce qui a pour conséquence que ses vecteurs propres, rangés par ordre croissant de valeur propre, "oscillent" de plus en plus (nombre de changements de signe dans les composantes strictement croissant).

Or, si $Y = \sum_k \gamma_k v_k$ et si ψ_k désignent les splines qui interpolent les vecteurs v_k, alors :

$$(2.8) \qquad y_0 = \sum_k \gamma_k \psi_k \quad \text{et} \quad y_\lambda = \sum_k (1 + \lambda \mu_k)^{-1} \gamma_k \psi_k .$$

On voit ainsi que le lissage par spline réalise un filtrage des hautes fréquences.

3 - REGRESSION NON-PARAMETRIQUE

3.1 - Modèle le plus courant

Les observations y_i s'écrivent $f(t_i) + \varepsilon_i$, où f désigne la fonction à estimer et ε_i une erreur aléatoire.

Les hypothèses sont doublement non paramétriques :

i) sur f : f est supposée lisse (smooth) se qui se traduit par l'appartenance de f à un espace fonctionnel H, non obligatoirement paramétré par un nombre fini de réels.

ii) sur les erreurs : $E(\varepsilon_i) = 0$, $E(\varepsilon_i\varepsilon_j) = \sigma^2\delta_{ij}$, mais il n'y a pas d'hypothèse sur la nature de la distribution.

3.2 - Estimateur spline conventionnel

Parallèlement au très populaire estimateur à noyau (cf. Collomb (1981,1985) et Härdle(1989) pour une revue), apparaît en Statistique vers 1975, l'estimateur spline présenté naturellement comme un estimateur des moindres carrés pénalisés. En effet, les moindres carrés paramétriques conduisent à une simple interpolation dès que l'espace fonctionnel est assez grand, et donc à une variance importante. En pénalisant convenablement le terme de moindre carrés comme en Ridge Regression, on se situe entre l'interpolation par fonctions splines (pénalisation quadratique) et la régression polynômiale. Le paramètre λ permet de doser biais et variance et de réaliser ainsi une meilleure erreur quadratique moyenne. Le rôle de la matrice $X\,({}^tXX)^{-1}\,{}^tX$ de la régression paramétrique est ici joué par la matrice A_λ appelée "influence matrix", "prediction matrix" ou "hat matrix".

3.3 - Vitesse de convergence

Pour évaluer les estimateurs splines, on peut calculer l'erreur quadratique moyenne

i)locale : $E[y_\lambda(t) - f(t)]^2 = R_\lambda(t)$.

ii)intégrée : $R_\lambda = \int_0^1 R_\lambda(t)\,dt$.

On étudie ensuite les vitesses de convergence de R_λ vers 0 lorsque n tend vers l'infini et la suite de points $\{t_i \ ; i = 1,...,n\}$ est dense dans [0 , 1], selon les hypothèses faites sur f, et pour des suites (λ_n) bien choisies (théoriquement ici et non à partir des données) de façon à optimiser ces vitesses.

Alors que dans un modèle paramétrique (hypothèses "draconiennes" sur f), le risque intégré d'un estimateur paramétrique adapté au modèle (i.e. sans biais) tend vers 0 comme 1/n, Speckman (1984) montre que la vitesse optimale (au sens du choix de λ_n) atteinte par les D^msplines est de l'ordre de $n^{-(2m/(2m+1))}$ pour f dans l'espace $W^{(m)}$. Cette vitesse est optimale dans la classe des estimateurs linéaires lorsque f appartient à $W^{(m)}$. Cox(1983) obtient des résultats analogues pour les splines robustes.

Speckman (1985) et Nussbaum (1985) recherchent des estimateurs minimax, parmi les estimateurs linéaires pour le premier et quelconques pour le second, pour des classes de fonctions du type $\{h \in W^{(m)}, \| D^m h \|^2 \leq P \}$. Ils montrent que ce sont des estimateurs splines qui ne sont pas du type conventionnel, mais qui correspondent à d'autres formes de filtres (cf. 2.8). Nussbaum donne une expression explicite de la constante dans l'équivalent asymptotique du risque minimax.

Avec les α-splines périodiques, Thomas-Agnan(1988b) montre que l'on peut atteindre des vitesses de l'ordre de $(\ln n)^{1/j} n^{-1}$ (j entier) dans des espaces plus petits que les $W^{(m)}_{per}$. Existe-t-il une classe de fonctions (non-paramétrique) que l'on puisse estimer avec un risque en 1/n ?

3.4 - Problèmes aux bords

Une étude du risque local montre que les termes dominants dans le risque intégré font intervenir le comportement de f aux extrémités de l'intervalle (cf. Rice et Rosenblatt(1983), Messer(1987)) . Ceci est dû au fait que si les t_i appartiennent à l'intervalle [0,1], la D^m-spline y_α satisfait aux conditions : $y_\alpha^{(m)}(1) = y_\alpha^{(m)}(0) = 0$.

Lorsque la vraie fonction f ne satisfait pas à ces mêmes conditions, la qualité de l'approximation et, par conséquent, du biais est moins bonne au voisinage de ces deux points qu'à l'intérieur de l'intervalle. Le même phénomène se produit pour les estimateurs à noyaux s'ils ne sont pas convenablement modifiés aux bords

3.5 - Equivalence avec la méthode des noyaux

Par linéarité, on peut écrire l'estimateur spline y_λ sous la forme :

$$y_\lambda(t) = \frac{1}{n} \sum_{i=1}^{n} y_i \, G_\lambda(t_i,t).$$

Plusieurs auteurs comparent les poids $G_\lambda(t_i,t)$ à ceux d'un estimateur à noyau qui sont $K[(t - t_i)/b]/b$ pour le noyau K et la largeur de bande b. D'autres auteurs envisagent une comparaison asymptotique mais l'on notera que la notion d'équivalence asymptotique est prise dans des sens différents selon les articles.

Cogburn et Davis(1974) montrent que, dans le cas périodique équidistribué ($t_i = i/n$), les poids $G_\lambda(t_i,t)$ d'un estimateur D^m-spline sont asymptotiquement comparables à ceux du noyau $\mathcal{K}_m$ dont la transformée de Fourier est $(1 + \omega^{2m})^{-1}$ et si l'on utilise la largeur de bande $b = \lambda^{1/2m}$.

Silverman(1984a) obtient un résultat similaire dans le cas des D^msplines non nécessairement périodiques ni à nœuds équidistribués, avec le même noyau $\mathcal{K}_m$ et une largeur de bande adaptée à la densité locale $\rho(t)$ des points t_i : $b = \lambda^{1/2m} \rho(t)^{-1/2m}$
Messer(1987) approfondit cette relation en montrant en particulier la différence de comportement aux extrémités de l'intervalle des deux estimateurs. Avec les α-splines périodiques, on élargit le résultat de Cogburn et Davis(1974) à une famille plus large de noyaux, comprenant par exemple le noyau Gaussien (voir Thomas-Agnan(1988b)).

Sur le plan pratique, Härdle(1989) compare les estimations faites sur les mêmes exemples par une méthode du noyau et par un lissage spline. Les résultats montrent une solution spline plus lisse mais obtenue avec une procédure numérique plus complexe. D'autre part, un estimateur spline peut prendre des valeurs négatives indésirables (comme le noyau équivalent) mais permet d'obtenir facilement les estimations des dérivées successives.

4 - PREDICTION, FILTRAGE

4.1 - Modèle de base

Les observations y_i se décomposent ici en somme d'un signal et d'un bruit. Le signal est une réalisation d'un processus aléatoire $Y(t_i)$ où $Y(t)$ est lui-même la somme d'une partie paramétrique $\Sigma\, \theta_k\, z_k(t)$, avec θ_k des paramètres inconnus et $z_k(.)$ des fonctions connues, et d'un processus du second ordre $X(t)$ de moyenne nulle, de covariance $R(s,t)$ (fonction connue). Le bruit suit le même modèle que précédemment et, de plus, il est supposé non corrélé avec le signal. Les problèmes de prédiction et filtrage consistent à estimer $Y(t)$ au vu de $y_1, y_2, \ldots, y_n$.

Dans certains cas, différents auteurs ont mis en évidence le fait que l'estimateur sans biais de variance minimum (ESBVM) de $Y(t)$ dans ce modèle est une fonction spline, d'interpolation si le bruit est nul, et de lissage sinon. Citons tout d'abord l'article sur l'équivalence formelle de Matheron(1981) qui se place dans un cadre abstrait. Pour les splines "concrètes", on peut préciser les types de processus concernés :

a - Kimeldorf et Wahba(1970a)

S'il existe un opérateur différentiel linéaire L d'ordre m tel que d'une part son noyau coïncide avec l'espace vectoriel engendré par les fonctions $z_i(.)$, et que d'autre part il soit lié à la structure de covariance $R(.,.)$ par l'intermédiaire de sa fonction de Green $G(.,.)$ et de la formule :

$$R(s,t) = \int_0^1 G(s,r)\ \ G(t,r)\, dr,$$

alors l'ESBVM est une fonction spline Tchebycheffienne et le paramètre de lissage est $\lambda = \sigma^2/n$. On trouve une représentation du processus $Y(.)$ dans ce cas en fonction du mouvement brownien dans Kimeldorf et Wahba (1971). Si de plus (cas a2), L est à coefficients constants et son polynôme caractéristique P_L n'a pas de racine réelle, alors le processus $X(.)$ est stationnaire de densité spectrale $|P_L|^{-2}$ (cf. Kimeldorf et Wahba(1970b)).

b - Salkaukas(1982), Dubrule(1983), Watson(1983)

Les hypothèses sur $X(.)$ sont légèrement différentes : c'est une fonction aléatoire intrinsèque à différences d'ordre (m-1) stationnaires, de variogramme proportionnel à

$\|t\|^{2m-1}$, où t est éventuellement multidimensionnel (cf. Matheron(1973)). Si, de plus, la partie paramétrique est engendrée par des polynômes, alors ces trois auteurs mettent en évidence à divers degrés de généralité que l'ESBVM est une D^m-spline.

c - Thomas-Agnan(1988a)

Si la partie paramétrique est nulle et si X(.) est stationnaire de densité spectrale $1/\alpha^2$ (intégrable par rapport à la mesure de Lebesgue), alors l'ESBVM est une α-spline (non périodique) et $\lambda = \sigma^2/n$ (ce qui généralise le cas a2).

d - Thomas-Agnan(1988b)

Dans le cadre du b, mais en dimension 1 et pour un variogramme combinaison linéaire de termes de la forme $|t|^{2k-1}$ avec certaines contraintes sur les coefficients ("Polynomial Generalized Covariance" dans Matheron (1973)), l'ESBVM est une α-spline non périodique pour une fonction α telle que α^2 soit rationnelle. Ceci recouvre le cas des processus du second ordre à différences d'ordre quelconque stationnaires.

On remarque que, dans tous ces cas d'équivalence concrète, il y a un lien entre la moyenne du signal et sa structure de covariance.

4.2 - Modèles Bayésiens

En ajoutant une hypothèse de distribution gaussienne pour le signal et le bruit, on peut envisager l'ESBVM comme une espérance conditionnelle de Y(t) sachant les données y_i, et ainsi, le problème de prédiction de processus rejoint alors la régression non paramétrique dans un cadre Bayésien.

a - Silverman(1985)

Silverman s'appuie sur la décomposition des données et de la spline dans la base des vecteurs propres de Ω pour mettre une distribution a priori (partiellement impropre) sur les coefficients γ_k. On peut qualifier ce modèle d'artificiel dans le sens où les éléments propres de Ω ne sont pas connus en général.

b - Kimeldorf et Wahba(1970b)

Dans le cas a2 du 4.1, Kimeldorf et Wahba supposent que Y(.) est un processus autorégressif gaussien d'ordre m, de moyenne nulle, et montrent que l'estimateur Bayésien est une fonction spline Tchebycheffienne. On remarque que, bien que cet

estimateur appartienne à un espace de fonctions lisses H_L, la probabilité a priori qu'une réalisation de ce processus n'appartienne pas à cet espace est égale à 1.

c - Wahba(1978)

Dans le cadre du modèle 4.1 a, avec $L = D^m$, l'auteur considère comme distribution a priori sur Y(.) celle du mouvement brownien (m-1)-intégré. Dans ces conditions, l'estimateur Bayésien est une D^mspline si les paramètres θ_i sont des constantes inconnues. Si on met également une distribution a priori sur les θ_i, normale de variance ξ^2, alors l'estimateur spline apparaît comme la limite, lorsque ξ tend vers l'infini, des estimateurs Bayésiens obtenus à ξ fixé (on parle alors de distribution a priori diffuse sur θ).

5 - ESTIMATION DU PARAMETRE DE LISSAGE OPTIMAL

Le choix d'une valeur correcte pour le paramètre de lissage qui contrôle le compromis entre régularité de la solution et fidélité aux données est fondamental pour la qualité de l'estimation dans les méthodes faisant appel à un lissage par fonctions splines. L'objectif est la recherche de la valeur λ_{opt} du paramètre qui minimise la fonction de perte $P(\lambda)$ exprimant l'erreur quadratique moyenne entre les mesures sans erreurs des données Bf et leur estimation Y_λ :

$$(5.1) \qquad \lambda_{opt} = \text{Arg} \min_{\lambda>0} [P(\lambda)] \quad \text{où} \quad P(\lambda) = \frac{1}{n} \| Y_\lambda - Bf \|^2.$$

Comme f est en pratique inconnue, différentes stratégies ont été proposées pour tenter d'estimer la valeur optimale λ_{opt} par la minimisation d'un critère spécifique. Celles-ci sont théoriquement applicables à l'expression la plus générale du problème mais des restrictions apparaissent pour certains algorithmes de calcul (cf. §6). Enfin la minimisation mise en œuvre n'est pas nécessairement convexe et peut ainsi conduire à une solution locale (cf. Wendelberger(1987)).

La première approche fut proposée par Reinsch(1967,1971) qui suggère de choisir λ de sorte que le résidu $(I - A_\lambda)Y$ soit une v.a. centrée et de variance (supposée connue ou estimée) σ^2. Mais cette démarche, trop dépendante de la qualité de l'estimation qui peut être faite de σ^2, conduit à surlisser les données (cf. Wahba(1975)).

5.1 - C_P de Mallows

Par analogie avec les modèles paramétriques, un critère est construit à partir du C_P de Mallows(1973). On désigne par $R(\lambda)$ le risque:

$$R(\lambda) = E[P(\lambda)] = \frac{1}{n}\|EY_\lambda - Bf\|^2 + \frac{\sigma^2}{n} tr\,(A_\lambda^2) \tag{5.2}$$

dont :

$$C_P(\lambda) = \frac{1}{n}\|(I - A_\lambda)Y\|^2 + \frac{\sigma^2}{n}(2\,trA_\lambda - n) \tag{5.3}$$

est un estimateur sans biais. Si σ^2 est connue ou estimée, la minimisation de C_P fournit une estimation de λ_{opt}. Dans l'analogie avec les modèles paramétriques, $tr(A_\lambda)$ joue le rôle du nombre de paramètres à estimer.

Les critères suivants ne nécessitent pas la connaissance ou l'estimation de σ^2.

5.2 - Maximum de Vraisemblance

Sous des hypothèses de normalité, Wecker et Ansley(1983), Wahba(1985), Kohn et Ansley(1987) construisent des estimateurs du maximum de vraisemblance de λ_{opt} fondés sur le modèle bayésien de Wahba(1978) (cf.§4.). Ainsi, Wahba(1985) propose de résoudre:

$$\lambda_{mv} = Arg \min_{\lambda>0} (MV(\lambda)) \quad avec \quad MV(\lambda) = \frac{{}^tY(I - A_\lambda)Y}{[det^+(I - A_\lambda)]^{1/(n-m)}} \tag{5.4}$$

où $det^+(I-A_\lambda)$ désigne le produit des (n-m) valeurs propres non nulles de $I-A_\lambda$. Kohn et Ansley(1987) montrent que λ_{mv} est un estimateur du maximum de vraisemblance marginal de λ_{opt}, fournissent simultanément un estimateur $\widehat{\sigma}^2(\lambda_{mv})$ de σ^2 et présentent un algorithme performant (cf. §4).

5.3 - Validation croisée

La méthode la plus répandue pour estimer λ_{opt} à partir des données avec σ^2 inconnue est une adaptation de la validation croisée (CV) décrite par Allen(1971) puis Stone(1974).

Le principe général consiste à minimiser la moyenne des erreurs quadratiques (la fonction PRESS de Allen) entre une observation y_k et sa valeur estimée par le prédicteur $y_\lambda^{(k)}$ construit sans cette $k^{\text{ième}}$-observation :

$$(5.5) \qquad CV(\lambda) = \frac{1}{n}\|Y_\lambda^{(.)} - Y\|^2 \quad \text{où} \quad Y_\lambda^{(.)} = {}^t[y_\lambda^{(1)}, \ldots, y_\lambda^{(n)}].$$

En exprimant $Y_\lambda^{(.)}$ en fonction de Y (cf. Wahba(1977)), on montre que (5.5) devient :

$$(5.6) \qquad CV(\lambda) = \frac{1}{n}{}^t[(I - A_\lambda)Y]\ \mathrm{Diag}\left(\frac{1}{1 - a_{ii}}\right)^2 (I - A_\lambda)Y$$

où les $\{a_{ii}\ ;\ i=1,\ldots,n\}$ désignent les termes diagonaux de A_λ et $\mathrm{Diag}(\theta_i)$ la matrice diagonale construite à partir du vecteur $(\theta_i)_{1\le i\le n}$.

La validation croisée généralisée (GCV), présentée par Wahba et Wold(1975,a,b), Wahba(1977), Craven et Wahba(1979), est obtenue en introduisant des pondérations dans (5.5) :

$$(5.7) \qquad GCV(\lambda) = \frac{1}{n}{}^t(Y_\lambda^{(.)} - Y)\ \mathrm{Diag}(w_i)\ (Y_\lambda^{(.)} - Y).$$

Par des justifications :

intuitives (Golub, Heath et Wahba(1979) proposent de rendre le critère invariant par rotation de la base),

heuristiques (ça marche sur des simulations) et

théoriques (le bon comportement asymptotique, cf.§ 5.4),

les pondérations sont définies par :

$$(5.8) \qquad w_i = \left(\frac{1 - a_{ii}}{\frac{1}{n}\mathrm{tr}\,(I - A_\lambda)}\right)^2.$$

Ceci revient à approcher dans (5.6) les termes diagonaux de A_λ par leur valeur moyenne: $\frac{1}{n}\mathrm{tr}(I - A_\lambda)$. Le problème s'écrit alors:

$$(5.9) \qquad \lambda_{gcv} = \mathrm{Arg}\ \min_{\lambda>0}\ (GCV(\lambda)) \quad \text{avec} \quad GCV(\lambda) = \frac{\frac{1}{n}\|(I - A_\lambda)Y\|^2}{\left(\frac{1}{n}\mathrm{tr}\,(I - A_\lambda)\right)^2}$$

Dans le cas des splines périodiques et si les nœuds sont équidistants l'expression (5.9) est exacte ; elle redonne la fonction PRESS de Allen car les matrices Ω et donc A_λ sont circulantes.

5.4 - Comportement asymptotique

L'une des justifications théoriques essentielles des critères précédents est leur optimalité asymptotique. Elle n'est pas obtenue par la convergence de λ vers λ_{opt}, car λ doit tendre vers 0 lorsque n croît pour assurer la convergence de l'estimation de f, mais par l'étude de la limite de $J_n = \frac{R(\hat{\lambda})}{R(\lambda_{Eopt})}$ où $\hat{\lambda}$ minimise E[GCV(λ)] ou E[ML(λ)] et λ_{Eopt} minimise R(λ).

Ces problèmes ont successivement été conjecturés puis résolus dans différents contextes principalement par Wahba(1975), Craven et Wahba(1979), Utreras(1983), Wahba(1985), Utreras(1987).

Dans le cas de la Validation Croisée et, moyennant une hypothèse technique sur la répartition asymptotique des nœuds, on est assuré que $J_n = 1+o(1)$ dans le cadre de définition des splines le plus général (§ 2.1, Wahba, 1985). De plus, Utreras(1987) montre que pour les D^m splines sur un ouvert de $\mathbb{R}^n$ vérifiant certaines hypothèses, l'estimateur de Validation Croisée est compatible avec la convergence de y_λ vers f. C'est-à-dire que :

$$\lim_{n\to\infty} E[\|f - y_{\lambda_{gcv}}\|^2] = 0$$

car λ_{gcv} tend vers 0 à une vitesse compatible avec celle requise pour la convergence de y_λ mais sans atteindre la vitesse optimale.

En comparant les méthodes de validation croisée et de maximum de vraisemblance, Wahba(1985) prouve que, si f est lisse, λ_{gcv} est asymptotiquement plus intéressant que λ_{ml} car $R(\lambda_{Egcv})$ converge plus vite que $R(\lambda_{Eml})$. D'autre part, si f est bruitée, les méthodes ont des comportements asymptotiquement similaires. Enfin, des simulations pour différentes valeurs de n et σ^2 semblent montrer que la validation croisée est "globalement plus performante" que le maximum de vraisemblance mais cette opinion n'est pas partagée par Kohn et Ansley(1987).

Un résultat asymptotique plus fort a été établi par Li(1986) pour la ridge regression mais aussi pour le lissage par des D^m-splines dans le cas où les nœuds sont équidistants et f différente d'un polynôme de degré inférieur ou égal à (m-1). Sous ces conditions, l'estimateur $\hat{\lambda} = \lambda_{C_p}$ qui minimise le C_p de Mallows ou $\hat{\lambda} = \lambda_{gcv}$ de la validation croisée vérifient:

$$\frac{P(\hat{\lambda})}{P(\lambda_{opt})} \rightarrow 1 \text{ en probabilité.}$$

6 - PROBLEMES ALGORITHMIQUES ET NUMERIQUES

6.1 - Le lissage

Le calcul des valeurs lissées Y_λ ne nécessite pas l'explicitation de la matrice d'influence A_λ ; l'algorithme le plus utilisé est celui implémenté dans la bibliothèque IMSL, c'est la méthode de projection (cf. Reinsch (1967,1971), Anselone et Laurent(1968)) qui conduit à résoudre un système d'équations linéaires à (n-m) inconnues dont la matrice bande (2m+1)-diagonale est symétrique, définie positive. D'autres algorithmes ont été proposés par la suite afin, non pas d'améliorer la vitesse qui reste de l'ordre de O(n) (seules sont dénombrées les multiplications et divisions de réels), mais la stabilité en cas de proximité de certains nœuds.

Peuvent être citées : la méthode du raccordement de Paihua(1978), l'utilisation d'une base de B-splines par De Boor(1978) et l'approche de Ansley et Kohn(1985) ou Kohn et Ansley(1987) qui considère le lissage comme l'espérance conditionnelle relative à un processus stochastique. Cette dernière permet également d'évaluer des intervalles de confiance de l'estimation de la fonction et de ses dérivées.

Dans le cas plus général où la matrice B n'est pas l'identité, c'est-à-dire hors du cadre de la simple évaluation, l'algorithme devient la résolution d'un problème variationnel à l'aide, par exemple, de la méthode du gradient conjugué.

6.2 - Estimation de la valeur optimale du paramètre

La recherche du paramètre optimal pose plus de problèmes algorithmiques. L'optimisation par Validation Croisée, qui est la méthode la plus fréquemment utilisée, nécessite l'évaluation de GCV(λ) (5.9) pour, en pratique, une vingtaine de valeurs λ. Le numérateur (le résidu) est obtenu, avec le même type d'algorithmes que ci-dessus, en un

nombre d'opérations de l'ordre de O(n) mais l'évaluation du dénominateur nécessite le calcul de la trace de A_λ qui s'avère beaucoup plus coûteux dans le cas général. La procédure implémentée dans IMSL requiert de l'ordre de $O(n^3)$ opérations et $O(n^2)$ emplacements en mémoire, elle devient rapidement impraticable sauf dans le cas particulier de la Ridge Regression avec $\Omega = I$ où le coût est en O(n). Aussi plusieurs auteurs ont proposé des méthodes moins coûteuses en temps et en mémoire fondées sur une approximation de $tr(A_\lambda)$.

Pour la validation croisée appliquée aux splines polynômiales (D^m-splines), Utreras(1980) calcule une approximation des valeurs propres de A_λ (donc de la trace) par une approximation de celles de Ω. Si les noeuds sont équidistants et les pondérations identiques, le temps de calcul est en O(n) et dans le cas général (cf. Utreras(1981b)), il est en $O(n^2)$. L'erreur décroît en 1/n. Le résultat pour des nœuds quelconques est amélioré par Silverman(1984) qui obtient encore un algorithme en O(n), puis par Hutchinson et Hoog(1985) qui exploitent la structure bande de la matrice du système linéaire obtenu pour chaque valeur de λ afin d'aboutir en $O(m^2n)$ opérations à un résultat exact.

Girard(1987) propose un algorithme de type "Monte-Carlo" élémentaire pour estimer la trace, utilisable dans le cas le plus général de la validation croisée et dont la précision ($O(n^{-1/2})$) est satisfaisante pour de grandes valeurs de n. La démarche consiste à générer un vecteur pseudo-aléatoire ε de taille n et de covariance la matrice identité (bruit blanc) puis à calculer son lissage $\varepsilon_\lambda = A_\lambda\varepsilon$ (en O(n) opérations) afin d'estimer $n^{-1}\ tr(I - A_\lambda)$ par le produit scalaire $n^{-1}\ {}^t\varepsilon(\varepsilon - \varepsilon_\lambda)$. Si n est petit, l'estimation peut être améliorée en effectuant une moyenne sur plusieurs tirages indépendants. Cependant, il faut se méfier des capacités des générateurs de nombres pseudo-aléatoires sur les micro-ordinateurs.

Ces différents auteurs suggèrent d'utiliser la méthode de la Section Dorée pour optimiser alors $GCV(\lambda)$ afin d'éviter, dans la mesure du possible, la présence éventuelle de minima locaux.

Enfin, seuls Kohn et Ansley(1987) présentent un algorithme efficace en O(n) opérations pour évaluer l'estimateur optimal du maximum de vraisemblance appliqué dans le cadre général du modèle stochastique.

7 - AUTRES TYPES D'UTILISATION

La présentation de l'introduction fait implicitement référence à une situation visant à éliminer des erreurs ou bruits additifs et correspondant à l'utilisation "naturelle" des fonctions splines de lissage. Néanmoins, cet outil, ou sa restriction au cas de l'interpolation, est largement utilisé en association avec des objectifs très variés dont voici quelques exemples.

7.1 - Estimation de la densité

L'estimation non-paramétrique de la densité (Prakasa Rao(1983), Bosq et Lecoutre(1983)) est un des domaines dans lequel se sont développées les premières applications des splines en Statistique. Boneva, Kendall et Stefanov(1971) proposent des interpolations splines des histogrammes (les histosplines). Exploitant les propriétés de dérivabilité des splines, Lii(1978), Berlinet(1979,1980,1981), Delabroye(1980) dérivent l'interpolation spline de la fonction de répartition pour estimer la densité. Les études asymptotiques (Wahba(1975b), Berlinet(1980,1981)) montrent que les estimateurs obtenus par interpolation spline ont des propriétés comparables à celles d'estimations dérivées d'autres techniques. En revanche, le comportement des queues de ces estimateurs semble moins favorable (Lii(1978)) et dépend des conditions aux bornes (Berlinet(1980)).

Indirectement, certaines méthodes admettent pour solutions des splines ; Wahba(1981) avec certaines classes de fonctions orthogonales ou De Montricher, Tapia, et Thompson(1975) recherchent un estimateur du maximum de vraisemblance pénalisé sous les contraintes que la solution soit positive, de somme unité et appartienne à une variété de fonctions régulières (inclûse dans $W^{(m)}$).

7.2 - Moyenne et analyse en composantes principales d'un processus

Dans le cas où l'on dispose de plusieurs observations ou réplications y_{ij} des mesures effectuées sur p "individus", les données se présentent sous la forme d'un tableau (pxn) et sont considérées comme les réalisations d'un processus $X(\omega,t)$, supposé du second ordre, sur un échantillon de taille p.

Besse et Ramsay(1986) interpolent les trajectoires de ce processus par des L-splines pour en plonger l'acp (analyse en composantes principales) dans des espaces fonctionnels (de Sobolev) qui induisent, en dimension finie, des métriques permettant de

prendre en compte des informations (variations, courbures) habituellement négligées par la métrique canonique.

D'autres travaux sur les processus se situent dans le cadre général du lissage en considérant le modèle : $y_j(t_i) = X(\omega_j,t_i) + \varepsilon_{ij}$. Les erreurs sont supposées décorrélées ou de matrice de covariance connue mais aucune hypothèse n'est faite sur leurs lois.

Biritxinaga(1987) pose le problème de l'estimation de l'espérance m(t) du processus X(t). Différents estimateurs, construits à partir des moyennes empiriques $\overline{X}(t_i)$, sont comparés : l'interpolation L-spline, le lissage par une L-spline avec les nœuds pondérés par les variances empiriques, le lissage par des splines sur un convexe c'est-à-dire passant par des intervalles de confiance définis autour des moyennes empiriques et dont la longueur dépend de la variance empirique.

Besse(1989) propose d'estimer les trajectoires $X(\omega,.)$ du processus en associant une technique de lissage et une réduction de dimension par une acp qui, dans le cas d'un processus, revient à une décomposition de Karhunen-Loève. Les trajectoires sont estimées (reconstruites) dans la base des q premiers vecteurs principaux ($q \ll n$) de la matrice $A_\lambda V {}^t A_\lambda$ où A_λ est la matrice d'influence d'un lissage par L-spline et V la matrice de covariance empirique du processus. Ceci revient à calculer l'acp des données lissées. Un exemple construit avec des données simulées montre que, dans ce cas, la prise en compte de l'information relative à la matrice de covariance dans cette démarche permet d'améliorer nettement l'erreur quadratique moyenne, par rapport à un simple lissage ou une simple acp, à condition de choisir correctement les valeurs de λ et q.

7.3. - Utilisation "paramétrique" des splines

L'outil spline est utilisé d'une façon tout à fait différente par certains auteurs. L'objectif principal n'est plus d'interpoler ou lisser mais de construire un ensemble de fonctions lisses, éventuellement monotones, qui serviront à généraliser les méthodes linéaires classiques. Ces méthodes qui visent une estimation paramétrique de fonctions lisses (non-paramétriques) sont parfois appelées semi-paramétriques (Wahba(1986)).

a - Régression spline

Une spline (souvent cubique) est substituée à la classique droite de régression par Poirier(1973), Wold(1974), Smith(1979), ..., Agarwal et Studden(1980). Les coordonnées dans une base de D^m-splines (m fixé à priori) ainsi que le nombre des nœuds, bien inférieur au nombre des observations, et leurs positions, qui ne

correspondent pas forcément aux points d'observation, sont estimés en minimisant des moindres carrés.

b - Régression non linéaire

Dans les modèles du type $\psi(Y) = \Sigma_j \phi_j(X_j)$ Winsberg et Ramsay(1980) recherchent des transformations non linéaires des données lisses, monotones et optimales au sens de la vraisemblance sous des hypothèses convenables. Ces transformations sont construites en intégrant (I-spline) des combinaisons positives de B-splines. Dans le même contexte, des algorithmes "intensifs" ont été proposés par Friedman, Grosse et Stuetzle(1983) (Projection Pursuit Regression), Breiman et Friedman(1985) (Alternating Conditional Expectation), Stone(1985), Wahba(1986).

c. réduction de la dimension

Ces familles de splines sont également utilisées avec un objectif de réduction de la dimension d'un tableau en restreignant au mieux le rang de la matrice de covariance des $\phi_j(X_j)$ (Winsberg et Ramsay(1983)) toujours en optimisant une vraisemblance. De Leeuw et Rijckevorsel(1988), Rijckevorsel(1987) poursuivent le même objectif de réduction de la dimension par un algorithme de moindres carrés alternés sans contrainte de monotonie. Cette dernière approche est parfois présentée comme une généralisation du "codage disjonctif" (découpage en classes d'une variable quantitative) sous l'appelation de "codage flou" (cf. Besse et Vidal (1982) ou Rijckevorsel J.L.(1988) pour une bibliographie détaillée).

Comme tout article bibliographique celui-ci présente des lacunes. Il reflète probablement mieux les préoccupations actuelles de ses auteurs que l'importance "objective" de chacun des thèmes abordés dans une littérature volumineuse et disparate. Aussi serions-nous reconnaissants à tout lecteur qui voudrait bien nous communiquer les références d'articles "importants" ayant échappé à notre sélection. Enfin, cette "photo" prise en octobre 1988 est insensible à la dynamique des recherches en cours. Elle n'anticipe pas sur le développement rapide de certains sujets, comme le lissage multidimensionnel dans $\mathbb{R}^d$ dont l'importance va probablement croître en proportion inverse du coût des stations de travail graphiques.

REFERENCES

Agarwal G. et Studden W.J., 1980. Asymptotic integrated mean square error using least squares and bias minimizing splines, *Annals of Statistics* , vol.8, n°6, 1307-1325.

Allen D.M., 1971. The prediction sum of squares as a criterion for selecting variables, *Technical Report n°23, Dpt of Statistics*, University of Kentucky.

Anselone P.M., Laurent P.J., 1968. A general method for the construction of interpolating or smoothing spline functions, *Numerische Mathematik*, 12, 66-82.

Ansley C.F., Kohn R., 1985. Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions, *Annals of Statistics*, vol.13, n°4, 1286-1316.

Atteia M., 1965. Fonctions-splines généralisées, *Compte Rendu Académie des Sciences*, Paris, 261, 2149-2152.

Atteia M., 1965. Généralisation de la définition et des propriétés des "splines-fonctions", *Compte Rendu Académie des Sciences*, Paris, 260, 3550-3553.

Atteia M., 1968. Fonctions "splines" définies sur un ensemble convexe, *Numerische Mathematik* , 12, 192-210.

Berlinet A., 1979. Sur les méthodes splines en estimation de la densité, *Compte Rendu Académie des Sciences*, Paris, 288, A, 847-850.

Berlinet A., 1980. Espaces autoreproduisants et mesure empirique, méthodes splines en estimation fonctionnelle, thèse de spécialité, Lille.

Berlinet A., 1981. Convergence des estimateurs splines de la densité, *Publication de l'Institut de Statistique de l'Université de Paris*, vol. 23, n°2, 1-16.

Besse P., Ramsay J., 1986. Principal components analysis of sampled functions, *Psychometrika*, 51, 2, 285-311.

Besse P., Vidal C., 1982. Analyse des correspondances et codage par une probabilité de transition, *Statistique et Analyse des Données*, vol. 7, n°3, 1-25.

Besse P., 1989. Optimal metric in principal components analysis of longitudinal data, *Data analysis and Informatics V*, (Ed. E.Diday et al.), North Holland, Amsterdam.

Bosq D., Lecoutre J.P., 1987. *Théorie de l'estimation fonctionnelle*, Economica, Paris.

Biritxinaga E., 1987. Estimation spline de la moyenne d'une fonction aléatoire, thèse de spécialité, Pau.

Breiman L., Friedman J.H., 1985. Estimating optimal transformations for multiple regression and correlation, *Journal of the American Statistical Association*, 80, 580-619.

Cogburn R. et Davis H.T., 1974. Periodic splines and spectral density estimation, *Annals of Statistics* , vol.2, 1108-1126.

Collomb G., 1981. Estimation non-paramétrique de la régression: revue bibliographique, *International Statististical Review*, 49, 75-93.

Collomb G., 1985. Nonparametric regression: an up-to-date bibliography, *Statistics*, 16, 309-324.

Cox D.D., 1983. Asymptotics for M-type smoothing splines, *Annals of Statistics*, vol.11, n°2, 530-551.

Cox D.D., 1984. Multivariate smoothing spline functions, *SIAM Journal on Numerical Analysis*, vol. 21, n°4, 789-813.

Cox D.D., Koh E., Wahba G., Yandell B.S., 1988. Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models, *Annals of Statistics*, vol.16, n°1, 113-119.

Craven P., Wahba G., 1979. Smoothing noisy data with spline functions, *Numerische Mathematik* , 31, 377-403.

De Boor C.,1978. *A practical guide to splines*, Applied Mathematical. Science 27, Springer Verlag, New-York.

Delabroye M., 1980. Comparaison entre la méthode de l'histogramme et la méthode des fonctions splines cubiques, thèse de spécialité, Rouen.

De Leeuw J., van Rijckevorsel J.L.A., 1988. Beyond homogeneity analysis, in *Component and Correspondence Analysis*, ed. van Rijckevorsel J., De Leeuw J., Wiley, London.

De Montricher G.F., Tapia R.A., Thompson J.R., 1975. Nonparametric maximum likelihood estimation of probability densities by penalty function methods, *Annals of Statistics*, vol. 3, 1329-1348.

Demmler A., Reinsch C. ,1975. Oscillation matrices with spline smoothing, *Numerische Mathematik* , 24, 375-382.

Deny J., Lions J.L., 1954. Les espaces du type de Beppo-Levi, *Annales de l'Institut Fourier*, Grenoble, vol.5, 305-370.

Dubrule O., 1983. Two methods with different objectives: splines and kriging, *Journal of the International Association for Mathematical Geology* , 15, 245-257.

Duchon J., 1976. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces , *R.A.I.R.O. Analyse Numérique* , vol.10, n°12, 5-12

Friedman J.H., Grosse E., Stuetzle W., 1983. Multidimensional additive spline approximation, *SIAM Journal on Scientific and Statistical Computing*, 4, 291-301.

Girard D., 1987. Un algorithme simple et rapide pour la validation croisée généralisée sur des problèmes de grande taille, *Rapport de Recherche 669-M*, IMAG, Grenoble.

Girard D., 1988 Détection de discontinuités dans un signal (ou une image) par inf-convolution spline et validation croisée : un algorithme rapide non paramétré, *Rapport de Recherche 702-J-M*, IMAG, Grenoble.

Golub G.H., Heath M., Wahba G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, vol. 21, n°2, 215-223.

Härdle W., 1989. *Applied nonparametric regression,* à paraître.

Huber P., 1979. Robust smoothing, in *Robustness in Statistics*, ed. Launer et Wilkinson, Academic, New-York.

Hutchinson M.F., Hoog F.R., 1985. Smoothing noisy data whith spline functions, *Numerische Mathematik*, 47, 99-106.

Kimeldorf G., Wahba G., 1970a. Spline fonctions and stochastic processes, *Sankhya* Series A, 32, 173-180.

Kimeldorf G., Wahba G., 1970b. A correspondence between bayesian estimation on stochastic processes and smoothing by splines, *Annals of Mathematical Statistics* , vol.41, n°2, 495-502.

Kimeldorf G., Wahba G., 1971. Some results on tchebycheffian spline functions, *Journal of Mathematical Analysis and Applications* , 33, 82-95.

Kohn R., Ansley C.F., 1983. On the smoothness properties of the best linear unbiased estimate of a stochastic process observed with noise, *Annals of Statistics*, vol. 11, n°3, 1011-1017.

Kohn R., Ansley C.F., 1987. A new algorithm for spline smoothing based on smoothing a stochastic process, *SIAM Journal on Scientific and Statistical Computing*, vol. 8, n°1, 33-48.

Laurent P.J., 1972. *Approximation et optimisation*, Hermann, Paris.

Li K. C., 1986. Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing, *Annals of Statistics*, vol. 14, n°3, 1101-1112.

Lii K.S., 1978. A global measure of a spline density estimate, *Annals of Statistics*, vol. 6, 1138-1148.

Mallows C.L., 1973. Some comments on C_p, *Technometrics*, vol. 15, n°4, 661-675.

Matheron G., 1973. The intrinsic random functions and their applications, *Advances in Applied Probability*, 5, 439-468.

Matheron G., 1981. Splines and kriging : their formal equivalence, Syracuse University Geology Contribution 8, D.F. Merriam ed, Dept. Geology, Syracuse Univ., New-York.

Meinguet J., 1979. Multivariate interpolation at arbitrary points made simple, *Journal of Applied Mathematics and Physics* (ZAMP) 30, 292-304.

Messer K., 1987. A Comparison of a spline estimate to its equivalent kernel estimate, pré-publication.

Nashed M.Z., Wahba G., 1974. Generalized inverses in reproducing kernel spaces : an approach to regularization of linear operator equations, *SIAM Journal on Mathematical Analysis*, vol 5, n° 6, 974-987.

Nussbaum M., 1985. Spline smoothing in regression models and asymptotic efficiency in L^2, *Annals of Statistics*, vol. 13,n°3, 984-997.

Paihua L., 1978. Quelques méthodes numériques pour le calcul des fonctions splines à une et plusieurs variables, Thèse, Grenoble.

Poirier D.J., 1973. Piecewise regression using cubic splines, *Journal of the American Statistical Association*, 68, 515-524.

Prakasa Rao B.L., 1983. *Nonparametric fonctionnal estimation*, Academic Press, New-York.

Reinsch C., 1967. Smoothing by spline functions, *Numerische Mathematik*, 10,177-183.

Reinsch C., 1971. Smoothing by spline functions II, *Numerische Mathematik* , 16,451-454.

Rice J., Rosenblatt M., 1983. Smoothing by splines : regression, derivatives, and deconvolution, *Annals of Mathematical Statistics*, vol.11, n°1, 141-156.

Salkaukas K., 1982. Some relationships between surface splines and kriging, Multivariate Approximation Theory II, *International Series of Numerical Mathematics* 61, Birkhäuser.

Schoenberg J., 1964. Spline functions and the problem of graduation, *Proc. Nat. Acad. Sci.* U.S.A. 52, 947-950.

Schumaker L., 1981. *Spline functions : basic theory*, Wiley, New-York.

Schumaker L.L., Utreras F., 1988. Asymptotic properties of complete smoothing splines and applications, *SIAM Journal on Scientific and Statistical Computing*, vol. 9, n°1, 24-38.

Silverman B.W., 1984a. Spline smoothing : the equivalent variable kernel method, *Annals of Statistics* , vol.12, n°3, 898-916.

Silverman B.W., 1984b. A fast and efficient cross-validation method for smoothing parameter choice in spline regression, *Journal of the American Statistical Association*, vol. 79, n°387, 584-589.

Silverman B.W., 1985. Some aspects of the spline smoothing approach to nonparametric regression curve fitting , *Journal of the Royal Statistical Society*, B, vol.47, n°1, 1-52.

Smith P., 1979. Splines as a useful statistical tool, *American Statistician*, 33, 57-62.

Speckman P., 1984. The asymptotic integrated mean square error for smoothing noisy data by Splines , manuscrit non publié.

Speckman P., 1985. Spline smoothing and optimal rates of convergence in nonparametric regression models, *Annals of Statistics* , vol.13, n°3, 970-983.

Stone M., 1974. Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statististical Society*, B, vol. 36, n°2, 111-147.

Stone C.J. ,1985. Additive regression and other nonparametric models, *Annals of Statistics*, vol.13, 689-705.

Thomas-Agnan C., 1988a. A family of splines for nonparametric regression and their relationships with kriging, pré-publication.

Thomas-Agnan C., 1988b. Smoothing periodic curves by a method of regularization, pré-publication.

Utreras F., 1980. Sur le choix du paramètre d'ajustement dans le lissage par fonctions splines, *Numerische Mathematik,* 34, 15-28.

Utreras F., 1981a. On computing robust splines and applications, *SIAM Journal on Scientific and Statistical Computing,* vol. 2, n°2, 153-163.

Utreras F., 1981b. Optimal smoothing of noisy data using spline functions, *SIAM Journal on Scientific and Statistical Computing*, vol 3, n°2, 349-362.

Utreras F., 1983. Natural spline functions, their associated eigenvalue problem, *Numerische Mathematik* , 42, 107-117

Utreras F., 1987. On generalized cross-validation for multivariate smoothing spline functions, *SIAM Journal on Scientific and Statistical Computing,* vol. 8, n°4, 630-643.

Van Rooij P.L.J., Schurer F., 1974. A bibliography on spline functions, in *Tagung über Spline Funktionen*, ed. Böhmer et al., Bibliographisches Institut, Mannheim, 315-415.

Van Rijckevorsel J.L.A., 1987. *The application of horseshoes and fuzzy coding in multiple correspondence analysis*, DSWO press, Leiden.

Van Rijckevorsel J.L.A., 1988. Fuzzy coding and B-splines, in *Component and Correspondence Analysis*, ed. van Rijckevorsel J., De Leeuw J., Wiley, London.

Wahba G., 1975. Smoothing noisy data with spline functions, *Numerische Math.*, 24, 383-393.

Wahba G., 1977. Practical approximate solutions to linear operator equations when data are noisy, *SIAM Journal on Numerical Analysis*, vol. 14, n°4, 651-667.

Wahba G., 1978. Improper priors, spline smoothing and a problem of guarding against models errors in regression, *Journal of the Royal Statististical Society,* B, 3, 364-372.

Wahba G., 1981. Data-based optimal smoothing of orthogonal series density estimates, *Annals of Statistics*, vol. 9, n°1, 146-156.

Wahba G., 1985. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, *Annals of Statistics*, vol..13, n°4, 1378-1402.

Wahba G., 1986. Partial and interaction spline models for the semiparametric estimation of functions of several variables, *Proceedings of the 18th Symposium on the Interface*, Colorado, March 1986, ed T.J. Boardman, Colorado State University.

Wahba G., Wold S., 1975a. A completely automatic french curve: fitting spline functions by cross-validation, *Communications in Statistics*, vol. 4, n°1, 1-17.

Wahba G., Wold S., 1975b. Periodic splines for spectral density estimation : the use of cross validation for determining the degree of smoothing, *Communications in Statistics* , vol.4, n°2, 125-141

Watson G.S., 1983. Smoothing and interpolation by kriging and with splines, *Technical Report* , 241, Series 2, dept Stat, Princeton University.

Wecker W.E., Ansley C.F., 1983. The signal extraction approach to nonlinear regression and spline smoothing, *Journal of the American Statistical Association*,vol. 78, n°381, 81-89.

Wegman E.J., Wright I.W., 1983. Splines in Statistics, *Journal of the American Statistical Association*, vol.78, n°382, 351-365.

Wendelberger J.G., 1987. Multiple minima of the generalized cross-validation function: paint attribute data, General Motors Research Laboratories Publications, GMR 5767.

Whittaker E.T., 1923. On a new method of graduation, *Proc. Edinburgh Math. Soc.*, 41, 63-75.

Winsberg S., Ramsay J.O., 1980. Monotonic transformations to additivity using splines, *Biometrika*, 67, 3, 669-674.

Winsberg S., Ramsay J.O., 1983. Monotone spline transformations for dimension reduction, *Psychometrika*, 48, 575-595.

Wold S., 1974. Spline functions in data analysis, *Technometrics*, 16, 1-11.

Wright I.W., Wegman E.J., 1980. Isotonic, convex and related splines, *Annals of Statistics*, vol. 8, n°5, 1023-1035.