Chapitre 6

# Modélisation statistique de données fonctionnelles

## 6.1. Introduction

Ce chapitre est consacré à l'étude d'ensembles de données dans lesquels les $n$ individus ou observations ne sont plus considérées comme de simples vecteurs de $\mathbb{R}^p$ mais sont des courbes ou plus généralement des fonctions. Ces fonctions dépendent d'un indice, traditionnellement le temps $t$, évoluant dans un intervalle que l'on supposera être, sans perte de généralité, un intervalle $T = [a, b]$ de $\mathbb{R}$. En pratique, ces fonctions sont observées en des instants de discrétisation qui peuvent être équirépartis ou non, identiques ou non, pour chaque courbe. La figure 6.1 donne un exemple type représentant des cumuls mensuels de précipitation.

Depuis une vingtaine d'années, ce type de données se rencontre de plus en plus fréquemment avec l'automatisation et l'informatisation des procédures de mesure. En conséquence, la littérature consacrée à l'étude de données fonctionnelles s'est considérablement développée.

Historiquement, les premiers travaux peuvent être attribués à des météorologues ou encore des chimistes qui furent les premiers à être confrontés à ce type de données ou encore à des techniques de traitement du signal associant Analyse en Composantes Principales (ACP) et décomposition de Karhunen-Loeve. En France, Deville [DEV 74] introduit une ACP de courbe tandis que, Dauxois et Pousse [DAU 76]

---

Chapitre rédigé par Philippe BESSE et Hervé CARDOT.

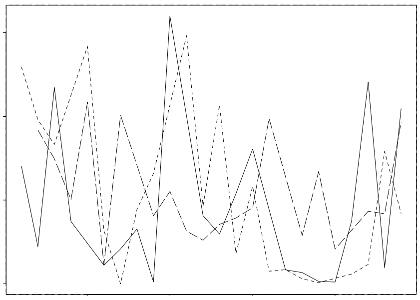

**Figure 6.1.** *Trois exemples de courbes décrivant la pluviométrie mensuelle durant 2 ans*

ont proposé un cadre synthétique généralisant l'analyse des données multidimensionnelles aux variables aléatoires hilbertiennes. La plupart des méthodes statistiques sont concernées et peuvent être déclinées dans un espace fonctionnel pour s'adapter à l'étude de courbes : analyse canonique [LEU 93], analyse Discriminante [HAS 95, DO 99], modèle linéaire [RAM 91, CAR 99a], modèle linéaire généralisé [MAR 99], régression inverse par tranches (SIR) [YAO 00]... La liste des références sur ce thème devient longue mais nous ne proposons pas de revue bibliographique dans ce chapitre et renvoyons à Ramsay et Silverman [RAM 97] pour un tour d'horizon plus exhaustif. Notons encore que l'étude de courbes peut nécessiter des transformations préalables (*curve registration*) afin de synchroniser et atténuer l'effet de certains évènements comme la date de la puberté qui intervient fortement dans des courbes de croissance. C'est obtenu par une transformation non linéaire des abscisses. Ce point, développé par exemple par Ramsay et Li [RAM 98] ou Kneip et coll. [KNE 00], n'est pas abordé dans ce chapitre.

L'adaptation de méthodes statistiques à des données fonctionnelles requiert un arsenal mathématique pouvant paraître sophistiqué voire rebutant. Certains de ces outils théoriques ne sont indispensables que pour aborder les aspects asymptotiques. Il est en effet important de pouvoir exhiber les analyses limites dans des espaces fonctionnels de dimension infinie lorsque le pas de discrétisation décroît et que la taille de l'échantillon croît indéfiniment. C'est le moyen de s'assurer de la stabilité des solutions proposées. Néanmoins, en pratique les données sont de toute façon discrétisées et les calculs réalisés matriciellement dans des espaces de dimension finie. D'autres outils, essentiellement issus de l'analyse numérique, sont alors indispensables pour rendre leur caractère fonctionnel aux observations discrétisées. Il s'agit principalement de techniques d'interpolation et de lissage (splines) ou de décomposition (Fourier, ondelettes) sur différents types de base. Les questions qui se posent alors sont les suivantes.

1) Dans quelles situations une approche fonctionnelle peut s'avérer plus efficace que celle vectorielle classique dans le cadre euclidien usuel $\mathbb{R}^p$ ?

2) Quels sont, en pratique, les compléments indispensables ?

Deux situations relativement fréquentes ont été identifiées comme répondant à la première question :

– lorsque les *courbes* ou fonctions observées sont très *régulières*, les variables issues de la discrétisation sont très corrélées deux à deux. Ceci a pour effet de masquer (effet « taille » trivial) l'essentiel des phénomènes d'intérêt ou de poser des problèmes de conditionnement de matrice. Dans le cadre de l'ACP, Besse et Ramsay [BES 86] proposent des pistes développées ensuite par Ramsay [RAM 96, RAM 00]),

– lorsque les données sont l'observation bruitée d'un phénomène que l'on peut supposer relativement régulier, il est important de faire intervenir une action de débruitage. Le problème soulevé est alors celui d'une coordination optimale entre une technique de débruitage et celle multidimensionnelle considérée.

Pour répondre à la deuxième question, nous avons choisi, dans ce chapitre, de ne donner que de courtes introductions aux cadres fonctionnels théoriques nécessaires aux études asymptotiques afin d'introduire les notations et la bibliographie. Ils peuvent être sautés en première lecture. Nous insistons en revanche sur la mise en œuvre matricielle de quelques méthodes en explicitant les outils numériques nécessaires. Ces explicitations matricielles sont données dans les situations les plus élémentaires correspondant au cas de données supposées régulières mais bruitées en se limitant aux techniques de lissage spline.

Trois types de méthodes sont abordés.

**ACP** La description statistique de courbes observées en présence de bruit conduit à une adaptation de l'analyse en composantes principales qui, dans ce cas, rejoint l'objectif de la régression non paramétrique en proposant une estimation *simultanée* de plusieurs courbes. Dans cet environnement fonctionnel, l'ACP doit incorporer des outils d'analyse numérique adaptés pour définir des approximations de courbes. Les fonctions splines d'interpolation et de lissage remplissent bien ce rôle mais d'autres techniques, comme la décomposition en ondelettes, auraient pu être utilisées.

**Modélisation** Elle concerne l'estimation de modèles pour lesquels les variables explicatives sont fonctionnelles. Plus généralement, l'introduction d'une fonction lien permet de considérer des modèles linéaires généralisées pour données fonctionnelles (régression logistique, ...). Diverses applications ont été proposées dans la littérature.

**Prévision** La prévision d'un processus à temps continu sur un intervalle de temps se traduit en termes de modèles pour données fonctionnelles par un changement d'écriture de ce processus. En effet, en le décrivant sous la forme d'un processus à temps discret fonctionnel (c'est-à-dire à valeurs dans un espace de fonctions), il est possible d'effectuer une prévision de ce processus en supposant qu'il est régi par des équations de type auto-régressifs. Il s'agit alors d'estimer l'opérateur d'auto-régression qui permet une prévision globale sur tout un intervalle de temps.

Pour simplifier la présentation de la méthodologie proposée, nous supposons dans ce chapitre que toutes les courbes sont observées selon le même plan de discrétisation c'est-à-dire aux mêmes instants. Dans le cas contraire, une adaptation a été proposée par Besse et coll. [BES 97] afin également de pouvoir prendre en compte des données manquantes. Celle-ci repose sur l'utilisation d'une approximation par splines hybrides associant B-splines [DEB 78, DIE 93]) et splines de lissage. Notons l'existence d'autres approches basées sur l'algorithme EM lorsqu'on observe uniquement des morceaux de courbes [RIC 01]. Enfin, tous les programmes[1] utilisés dans les exemples sont écrits en Splus [SPL 97].

## 6.2. Cadre fonctionnel

Cette section fournit un cadre mathématique général à ce chapitre. Un premier objectif est de définir les notations nécessaires à la manipulation de variables ou processus aléatoires à valeurs dans un espace fonctionnel. Incontournables pour des études asymptotiques, ces notions peuvent être survolées en première lecture. Le deuxième objectif est de définir des critères de *régularité* d'une fonction qui interviendront comme contraintes dans les optimisations ou termes de pénalisation. Ils s'exprimeront pratiquement par l'explicitation matricielle de normes ou semi-normes dans un espace euclidien de dimension finie.

### 6.2.1. *Variable aléatoire hilbertienne*

On considère une variable aléatoire $Z$ à valeurs dans l'espace de Hilbert supposé séparable $H$ muni de la tribu des boréliens. On note $\|x\|_H$ la norme dans cet espace. On suppose que $Z$ est du second ordre, c'est-à-dire qu'elle vérifie $\mathbb{E}\|Z\|_H^2 < \infty$. Sous cette hypothèse, $Z$ admet une espérance dans $H$ notée $\mathbb{E}(Z)$ et un opérateur de covariance compact $\Gamma$ admettant donc un spectre discret.

L'existence des moments de $Z$ et leur définition sont fournies par le théorème de Riesz ($H'$ désigne le dual topologique de $H$) :

$$\forall f \in H, \quad \langle \mathbb{E}(Z), f \rangle_H = \mathbb{E}[\langle f, Z \rangle_H],$$

1. Ils sont accessibles à partir de l'URL `www.inra.fr/bia/T/cardot/`

$$\forall (u,v) \in H' \times H', \ \langle \Gamma u, v \rangle_{H,H'} = \mathbb{E}\Big[ \langle Z - \mathbb{E}(Z), u \rangle_{H,H'} \langle Z - \mathbb{E}(Z), v \rangle_{H,H'} \Big].$$

L'opérateur de covariance s'écrit alors avec la notation de produit tensoriel :

$$\Gamma = \mathbb{E}[(Z - \mathbb{E}(Z)) \otimes (Z - \mathbb{E}(Z))].$$

Dans le cas particulier où $H = L^2(T)$, il est facile de vérifier que l'opérateur de covariance s'écrit sous la forme d'un opérateur intégral :

$$\forall f \in L^2(T), \forall t \in T, \quad \Gamma f(t) = \int_T \gamma(s,t) f(s) \, ds, \tag{6.1}$$

où $\gamma(s,t)$ est la fonction de covariance du processus à temps continu $Z(t)$ d'espérance $a(t)$ :

$$\gamma(s,t) = \mathbb{E}[(Z(t) - a(t))(Z(s) - a(s))], \quad \forall (s,t) \in T \times T. \tag{6.2}$$

Nous serons également amenés à considérer un processus à temps discret $(Z_i)_{i\in\mathbb{Z}}$ supposé du second ordre, auto-régressif d'ordre 1 et prenant ses valeurs dans un espace hilbertien : $(Z_i)_{i\in\mathbb{Z}}$ est dit ARH(1). Notons $\Gamma$ l'opérateur de covariance et $\Delta$ celui de covariance croisée du processus. Le processus étant supposé stationnaire, ces opérateurs ne dépendent pas de $i$ et sont définis par :

$$\Gamma = \mathbb{E}[(Z_0 - \mathbb{E}(Z_0)) \otimes (Z_0 - \mathbb{E}(Z_0))],$$
$$\Delta = \mathbb{E}[(Z_0 - \mathbb{E}(Z_0)) \otimes (Z_1 - \mathbb{E}(Z_1))],$$
$$\Delta^* = \mathbb{E}[(Z_1 - \mathbb{E}(Z_1)) \otimes (Z_0 - \mathbb{E}(Z_0))]$$

où la fonction $\mu$ représente la moyenne du processus $\mathbb{E}(Z_i)$. Ils vérifient :

$$\Delta^*(Z_i - \mu) = \Gamma(E(Z_{i+1}|Z_i, Z_{i-1}, \ldots) - \mu),$$
$$\Delta = \rho\Gamma$$

et possèdent un spectre discret.

### 6.2.2. *Condition de régularité*

Les différentes techniques proposées reposent sur la recherche de solutions régulières ou lisses au sens d'un critère faisant intervenir les normes des dérivées successives. Ce critère est couramment utilisé pour la définition et la construction des fonctions splines, il se définit comme une semi-norme dans un espace de Sobolev

$$W^m = \{z : z, z', \ldots, z^{(m-1)} \text{ absolument continues }, z^{(m)} \in L^2\}.$$

Pour toute fonction $z$ de $W^m$, sa régularité est contrôlée par la semi-norme :

$$\|z\|_m^2 = \|D^m z\|_{L^2(T)}^2 = \int_T (z^{(m)}(t))^2 dt. \tag{6.3}$$

Ce critère peut être généralisé à d'autres semi-normes équivalentes [WAH 90] en remplaçant l'opérateur $D^m$ par tout opérateur différentiel linéaire faisant au moins intervenir le même ordre $m$ de dérivation et conduisant ainsi à la définition de familles plus générales de splines dites de Tchebicheff.

### 6.2.3. *Splines de lissage*

L'estimation non-paramétrique par lissage spline a donné lieu à une importante littérature : [WAH 90], [GRE 94] en fournissent par exemple des présentations détaillées. Plaçons-nous dans le cadre usuel du modèle de régression non paramétrique :

$$x_j = z(t_j) + \varepsilon_j \ ; \quad E(\varepsilon_j) = 0,\ E(\varepsilon_j \varepsilon_k) = \sigma^2 \delta_{kj},\ j,k = 1,...,p \qquad (6.4)$$
$$a \leq t_1 < t_2 < ... < t_p \leq b.$$

où $x_j$ est l'observation bruitée en $t_j$ de la fonction $z$ supposée régulière et lisse : $z \in W^m$.

L'estimation spline $\widehat{z}$ de la fonction $z$ est alors la solution du problème d'optimisation sous contrainte de régularité

$$\min_{z \in W^2} \left\{ \frac{1}{p} \sum_{j=1}^{p} (z(t_j) - x(t_j))^2;\ \|z\|_m^2 < \text{c}\ (\text{c}\ \in \mathbb{R}_+) \right\}. \qquad (6.5)$$

En introduisant un multiplicateur de Lagrange, ce problème d'optimisation est équivalent à :

$$\min_{z \in W^2} \left\{ \frac{1}{p} \sum_{j=1}^{p} (z(t_j) - x(t_j))^2 + \ell \left\| D^m z \right\|_{L^2}^2 \right\}. \qquad (6.6)$$

Le paramètre de lissage $\ell$, qui dépend directement de $c$, permet d'effectuer un arbitrage entre la fidélité aux données ($\ell \to 0$) et la régularité de la solution ($\ell \to +\infty$). En régression non paramétrique sa valeur est choisie en minimisant le critère de validation croisée généralisée (GCV, [WAH 90]).

La solution $\widehat{z}$ de ce problème est une fonction polynômiale par morceaux ; polynômes de degré (2m-1) entre deux nœuds $t_j$ et $t_{j+1}$, de degré $(m-1)$ aux extrémités entre $a$ et $t_1$ et entre $t_p$ et $b$. À la limite $(\ell = 0)$ la solution est la fonction d'interpolation spline passant par les données observées et minimisant le critère. À l'autre limite ($\ell$ infini), la solution est une régression polynômiale de degré $(m-1)$ rendant nulle la pénalisation.

La construction explicite dépend alors de la base choisie pour définir le sous-espace $S_p$ des fonctions splines et différentes solutions sont proposées dans la littérature. La plus simple, sur le plan théorique, consiste à utiliser les propriétés d'autoreproduction de l'espace de Sobolev et donc son noyau [WAH 90] comme base. Cette approche, adoptée aussi par Besse et Ramsay [BES 86], pose des problème numériques lorsque, en pratique, le nombre de nœuds $p$ est grand car elle conduit à l'inversion d'une matrice mal conditionnée. Dans la version simplifiée de ce chapitre, nous nous limitons aux fonctions splines cubiques dites naturelles, c'est-à-dire polynômiales de degré 3 par morceaux ($m = 2$) et dont les dérivées secondes et troisièmes s'annulent aux bornes $a$ et $b$. L'algorithme de Reisch [GRE 94] conduit alors à la résolution numérique d'un système d'équations tridiagonal par décomposition de Cholesky puis substitution en un nombre d'opérations qui croît linéairement avec $p$. Il ne pose alors plus de problèmes numériques.

Soit $\mathbf{Q}$ la matrice bande $(p \times (p-2))$ d'éléments

$$q_{j-1,j} = \frac{1}{t_j - t_{j-1}}, q_{j,j} = -\frac{1}{t_j - t_{j-1}} - \frac{1}{t_{j+1} - t_j}, q_{j+1,j} = -\frac{1}{t_{j+1} - t_j},$$

sur les diagonales (0 à l'extérieur) et $\mathbf{R}$ la matrice bande symétrique $(p-2) \times (p-2)$ d'éléments

$$r_{j,j} = \frac{1}{3}(t_{j+1} - t_{j-1}), r_{j+1,j} = r_{j,j+1} = \frac{1}{6}(t_{j+1} - tj).$$

Notons $\mathbf{M}$ la matrice associée au semi-produit scalaire de $W^2(T)$ induit sur $S_p$ : si $z$ désigne une fonction de $W^2(T)$ et $\mathbf{z}$ le vecteur de ses valeurs aux nœuds,

$$\int_T z''(t)^2 dt = \|z\|_2^2 = \mathbf{z}'\mathbf{M}\mathbf{z} = \|\mathbf{z}\|_{\mathbf{M}}^2 .$$

La matrice $\mathbf{M}$ est définie par $\mathbf{M} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}$.

Formellement, si le vecteur $\mathbf{x}$ contient les $p$ observations aux nœuds $t_j$, le lissage spline revient à calculer le vecteur

$$\widehat{\mathbf{z}} = \mathbf{A}_\ell \mathbf{x} \quad \text{avec} \quad \mathbf{A}_\ell = (\mathbf{I} + \ell\mathbf{M})^{-1} \tag{6.7}$$

où $\mathbf{A}_\ell$ est la matrice de lissage (ou hat matrix ou encore smoothing matrix). Enfin, la fonction $\widehat{z}$ est obtenue par simple interpolation spline aux valeurs du vecteur $\widehat{\mathbf{z}}$. On remarque que les solutions obtenues ne dépendent pas des positions des bornes $a$ et $b$ de l'intervalle $T$.

Une autre matrice de produit scalaire $\mathbf{N}$ est nécessaire, il s'agit de celle associant aux vecteurs $\mathbf{y}_1$ et $\mathbf{y}_2$ issus d'un même schéma de discrétisation le produit scalaire dans l'espace $L^2(T)$ entre leur interpolant spline $\widehat{y}_1$ et $\widehat{y}_2$ :

$$\mathbf{y}_1'\mathbf{N}\mathbf{y}_2 \;=\; \int_T \hat{y}_1(t)\hat{y}_2(t)\; dt. \tag{6.8}$$

D'une manière générale cette matrice s'obtient à l'aide des noyaux reproduisants associés aux fonctions splines [BES 86, RAM 91]). Il est également possible de l'approcher en utilisant une méthode de quadrature. On peut considérer par exemple $\mathbf{N} = \operatorname{diag}(w_1, \ldots, w_p)$ où $w_1 = (t_2 - t_1)/2, w_j = (t_{j+1} - t_{j-1})/2,\ j = 2, \ldots, p-1$ et $w_p = (t_p - t_{p-1})/2$. Le calcul est alors rapide, stable et généralement suffisamment précis.

## 6.3. Analyse en composantes principales

Nous nous intéressons dans cette section à la description et à l'estimation des réalisations de trajectoires $z_i$ d'un processus $Z$ ou, c'est équivalent, d'une variable aléatoire prenant ses valeurs dans un espace hilbertien. Avec les notations de la section précédente, nous considérons que la variable aléatoire $X$ constitue l'observation bruitée des trajectoires, supposées régulières, de la variable aléatoire $Z$. La figure 6.1 donne un exemple illustratif de telles données.

Chacune des $n$ réalisations ou trajectoires $z_i$, est donc supposée observée pour un nombre $p$ d'instants de discrétisation $t_1, \ldots, t_p$ de l'intervalle $T$, les mêmes pour chaque trajectoire. Cette mesure introduit des erreurs aléatoires indépendantes et identiquement distribuées de variance $\sigma^2$. La situation correspond donc à $n$ répétitions supposées indépendantes du modèle (6.4) auquel il faut ajouter l'hypothèse d'indépendance entre les différentes réalisations de $Z$ et le bruit : $\mathbb{E}(\boldsymbol{\epsilon}_i \mathbf{z}_{i'}') = 0$.

À ce niveau, il serait possible de considérer l'estimation des $n$ trajectoires de $Z$ comme $n$ problèmes classiques d'estimation non paramétrique de fonctions de régression. Néanmoins, intuitivement et c'est vérifié par des simulations [BES 97], il semble important de tenir compte du fait qu'il s'agit de l'estimation simultanée de $n$ réalisations d'un même processus et donc de tenir compte de la structure de covariance qu'il est possible d'estimer. Ceci est pris en compte en introduisant une contrainte supplémentaire issue de l'hypothèse que la variable $Z$ évolue dans un sous-ensemble de dimension finie de $W^2(T)$. Cela revient encore à écrire que ses trajectoires s'expriment comme combinaisons linéaires d'un nombre réduit $q$ de composantes. Ces composantes étant par ailleurs régulières du fait de la première hypothèse.

### 6.3.1. *Modèle et estimation*

Les observations de chacune des trajectoires sont rangées dans des vecteurs $\mathbf{x}_i = (x_i(t_1), \ldots, x_i(t_p))'$ de $\mathbb{R}^p$ et $A_q$ désigne un sous-espace affine de $\mathbb{R}^p$ de dimension

$q < p$. La situation impliquée par l'estimation simultanée de $n$ régressions non paramétriques sous une double contrainte de régularité et de dimension se résume alors par le modèle suivant :

$$\begin{array}{l} \mathbf{x}_i = \mathbf{z}_i + \boldsymbol{\epsilon}_i; \\ i = 1, \ldots, n \end{array} \quad \text{avec} \left\{ \begin{array}{l} \mathbb{E}(\boldsymbol{\epsilon}_i) = 0 \text{ et } \mathbb{E}(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i') = \sigma^2 \mathbf{I}, \\ \sigma \text{ inconnue, } (\sigma > 0) \\ \mathbf{x}_i \text{ indépendant de } \boldsymbol{\epsilon}_{i'},\ i' = 1, \ldots, n, \\ \mathbf{x}_i \in A_q \text{ p.s.} \quad \text{et} \quad \|\mathbf{x}_i\|_m^2 \leq c \text{ p.s..} \end{array} \right. \tag{6.9}$$

Ce modèle présente donc la particularité d'associer deux types de contraintes, la première, de dimension, conduisant à une définition de l'analyse en composantes principales (ACP) proposée par Caussinus [CAU 86], la deuxième, de régularité, habituelle en statistique fonctionnelle. Son estimation par les moindres carrés pondérés nous amène à résoudre un problème d'optimisation dans lequel la contrainte de régularité a été remplacée par un multiplicateur de Lagrange $\ell$ dépendant de $c$.

Avec les notations matricielles et en supposant que les observations sont pondérées par les éléments diagonaux $w_i$ de la matrice $\mathbf{D}$, il s'agit de résoudre :

$$\min_{\mathbf{z}_i, A_q} \left\{ \sum_{i=1}^{n} w_i \left( \|\mathbf{z}_i - \mathbf{x}_i\|_{\mathbf{I}}^2 + \ell \|\mathbf{z}_i\|_{\mathbf{M}}^2 \right); \mathbf{z} \in A_q,\ \dim A_q = q \right\} \tag{6.10}$$

Notons par $\overline{\mathbf{x}} = \sum_{i=1}^{n} w_i \mathbf{x}_i$ la moyenne des coordonnées et par $\overline{\mathbf{X}}$ la matrice des observations centrées c'est-à-dire dans un contexte d'études climatiques, la matrice des anomalies $(\mathbf{x}_i - \overline{\mathbf{x}})$ des observations par rapport à la moyenne annuelle ; $\mathbf{S}$ désigne la matrice de covariance empirique : $\mathbf{S} = \overline{\mathbf{X}}'\mathbf{D}\overline{\mathbf{X}}$.

PROPOSITION.– *La solution du problème 6.10 est donnée par :*

$$\widehat{\mathbf{z}}_i = \mathbf{A}_\ell^{1/2} \widehat{\mathbf{P}}_q \mathbf{A}_\ell^{1/2} \mathbf{x}_i + \mathbf{A}_\ell \overline{\mathbf{x}}\ ,\ i = 1, \ldots, n.$$

*La matrice* $\widehat{\mathbf{P}}_q = \mathbf{V}_q \mathbf{V}_q'$ *est la projection orthogonale sur le sous-espace* $\widehat{E}_q$ *engendré par les q vecteurs propres de la matrice*

$$\mathbf{A}_\ell^{1/2} \mathbf{S} \mathbf{A}_\ell^{1/2}.$$

*associés aux q plus grandes valeurs propres.* Les estimations lisses des trajectoires s'obtiennent alors par interpolation spline des valeurs contenues dans le vecteur $\widehat{z}_i$.

*Démonstration*

Notons $\mathbf{z}_i$ le vecteur de $\mathbb{R}^p$ contenant les valeurs de $z_i$ et $\overline{\mathbf{z}} = \sum_{i=1}^{n} w_i \mathbf{z}_i$. On définit la matrice centrée $\mathbf{Z}$ $(n \times p)$ dont les vecteurs lignes $(\mathbf{z}_i - \overline{\mathbf{z}}')$ sont contraints à appartenir au sous-espace vectoriel $E_q = A_q - \overline{\mathbf{z}}$. Cette contrainte est équivalente à imposer à la matrice $\mathbf{Z}$ d'être au plus de rang $q$.

Le critère à minimiser se décompose de la façon suivante :

$$\sum_{i=1}^n w_i \left(\|\mathbf{z}_i - \mathbf{x}_i\|_{\mathbf{I}}^2 + \ell \|\mathbf{z}_i\|_{\mathbf{M}}^2\right) = \sum_{i=1}^n w_i \|\mathbf{z}_i - \overline{\mathbf{z}} - (\mathbf{x}_i - \overline{\mathbf{x}})\|_{\mathbf{I}}^2$$
$$+\ell \sum_{i=1}^n w_i \|\mathbf{z}_i - \overline{\mathbf{z}}\|_{\mathbf{M}}^2 + \|\overline{\mathbf{x}} - \overline{\mathbf{z}}\|_{\mathbf{I}}^2 + \ell \|\overline{\mathbf{z}}\|_{\mathbf{M}}^2 .$$

Les deux derniers termes de cette expression conduisent à estimer $\overline{\mathbf{z}}$ par lissage spline de la moyenne empirique :

$$\widehat{\overline{\mathbf{z}}} = \mathbf{A}_\ell \overline{\mathbf{x}} \quad \text{donc} \quad \widehat{A}_q = \widehat{\overline{\mathbf{z}}} + \widehat{E}_q.$$

Les deux premiers termes nous amènent ensuite à résoudre :

$$\min_{\mathbf{Z}(n\times p)} \left\{ \left\|\mathbf{Z} - \overline{\mathbf{X}}\right\|_{\mathbf{I},\mathbf{D}}^2 + \ell \|\mathbf{Z}\|_{\mathbf{M},\mathbf{D}}^2 ;\ \text{rang}(\mathbf{Z}) = q,\ q < p \right\} \tag{6.11}$$

où $\|\mathbf{Z}\|_{\mathbf{M},\mathbf{D}}^2 = \text{tr}\mathbf{Z}'\mathbf{D}\mathbf{Z}\mathbf{M}$ désigne la norme euclidienne des matrices $(n \times p)$.

Notons $\widetilde{\overline{\mathbf{X}}} = \overline{\mathbf{X}}\mathbf{A}_\ell$ la matrice des lignes lissées de $\overline{\mathbf{X}}$, de sorte que

$$\begin{aligned}
\left\|\mathbf{Z} - \overline{\mathbf{X}}\right\|_{\mathbf{I},\mathbf{D}}^2 + \ell \|\mathbf{Z}\|_{\mathbf{M},\mathbf{D}}^2 &= \text{tr}\overline{\mathbf{X}}'\mathbf{D}\overline{\mathbf{X}} - 2\text{tr}\overline{\mathbf{X}}'\mathbf{D}\mathbf{Z} + \text{tr}\mathbf{Z}'\mathbf{D}\mathbf{Z}(\mathbf{I} + \ell\mathbf{M}) \\
&= \text{tr}\widetilde{\overline{\mathbf{X}}}'\mathbf{D}\widetilde{\overline{\mathbf{X}}}(\mathbf{I} + \ell\mathbf{M})^2 - 2\text{tr}\widetilde{\overline{\mathbf{X}}}'\mathbf{D}\mathbf{Z}(\mathbf{I} + \ell\mathbf{M}) + \text{tr}\mathbf{Z}'\mathbf{D}\mathbf{Z}(\mathbf{I} + \ell\mathbf{M}) \\
&= \left\|\widetilde{\overline{\mathbf{X}}} - \mathbf{Z}\right\|_{(\mathbf{I}+\ell\mathbf{M}),\mathbf{D}}^2 + \ell \left\|\widetilde{\overline{\mathbf{X}}}\right\|_{\mathbf{M},\mathbf{D}}^2 + \ell^2 \left\|\widetilde{\overline{\mathbf{X}}}\right\|_{\mathbf{M^2},\mathbf{D}}^2 .
\end{aligned}$$

Seul le premier terme de cette équation dépend de $\mathbf{Z}$. Par conséquent, la solution est la meilleure approximation de rang $q$ de la matrice $\widetilde{\overline{\mathbf{X}}}$. Elle est obtenue par la décomposition en valeurs singulières (DVS) de $\overline{\mathbf{X}}\mathbf{A}_\ell$ relativement aux métriques $\mathbf{A}_\ell^{-1}$ et $\mathbf{D}$ :

$$\widehat{\mathbf{Z}_q} = \widetilde{\mathbf{U}}_q \widetilde{\mathbf{L}}_q^{1/2} \widetilde{\mathbf{V}}_q',$$

$$\text{où} \begin{cases} \overline{\mathbf{X}}\mathbf{A}_\ell\overline{\mathbf{X}}'\mathbf{D}\widetilde{\mathbf{U}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{L}} & \text{et} \quad \widetilde{\mathbf{U}}'\mathbf{D}\widetilde{\mathbf{U}} = \mathbf{I}, \\ \mathbf{A}_\ell\overline{\mathbf{X}}'\mathbf{D}\mathbf{X}\widetilde{\mathbf{V}} = \widetilde{\mathbf{V}}\widetilde{\mathbf{L}} & \text{et} \quad \widetilde{\mathbf{V}}'\mathbf{A}_\ell^{-1}\widetilde{\mathbf{V}} = \mathbf{I}. \end{cases}$$

Cette décomposition en valeurs singulières généralisée est aussi déduite de celle de $\overline{\mathbf{X}}\mathbf{A}_\ell^{1/2}$ relativement à $\mathbf{I}$ et $\mathbf{D}$ :

$$\overline{\mathbf{X}}\mathbf{A}_\ell^{1/2} = \mathbf{U}\mathbf{L}^{1/2}\mathbf{V}'$$

$$\text{où} \begin{cases} \overline{\mathbf{X}}\mathbf{A}_\ell^{1/2}\mathbf{A}_\ell^{1/2}\overline{\mathbf{X}}'\mathbf{D}\mathbf{U} = \mathbf{U}\mathbf{L} & \text{et} \quad \mathbf{U}'\mathbf{D}\mathbf{U} = \mathbf{I}, \\ \mathbf{A}_\ell^{1/2}\overline{\mathbf{X}}'\mathbf{D}\overline{\mathbf{X}}\mathbf{A}_\ell^{1/2}\mathbf{V} = \mathbf{V}\mathbf{L} & \text{et} \quad \mathbf{V}'\mathbf{V} = \mathbf{I}. \end{cases}$$

On retrouve ensuite $\widetilde{\mathbf{L}} = \mathbf{L}$, $\widetilde{\mathbf{U}} = \mathbf{U}$ et $\widetilde{\mathbf{V}} = \mathbf{A}_\ell^{1/2}\mathbf{V}$. □

La décomposition en valeurs singulières de $(\mathbf{X}\mathbf{A}_\ell^{1/2}, \mathbf{I}, \mathbf{D})$ conduit à l'analyse spectrale de la matrice $\mathbf{A}_\ell^{1/2}\mathbf{S}\mathbf{A}_\ell^{1/2}$. Les trajectoires discrètes du processus sont projetées sur le sous-espace engendré par les vecteurs

$$\widetilde{\mathbf{v}}_i \;=\; \mathbf{A}_\ell^{1/2}\mathbf{v}_j, \quad j = 1, \ldots, q.$$

Les trajectoires discrètes estimées $\hat{\mathbf{z}}_i$ se décomposent de manière équivalente sur la base $\mathbf{A}_\ell^{-1}$-orthonormée des $\{\widetilde{\mathbf{v}}_j\}$ par projection des données transformées $\mathbf{A}_\ell \mathbf{x}_i$ :

$$\hat{\mathbf{z}}_i = \mathbf{A}_\ell \bar{\mathbf{x}} + \sum_{j=1}^{q} \langle \widetilde{\mathbf{v}}_j, \mathbf{A}_\ell \mathbf{x}_i \rangle_{\mathbf{A}_\ell^{-1}} \widetilde{\mathbf{v}}_j. \tag{6.12}$$

### 6.3.2. *Dimension et paramètre de lissage*

Cette méthode nécessite de régler les valeurs de deux paramètres : la dimension $q$ du sous-espace ainsi que celle du paramètre de lissage $\ell$. Ce choix doit être réalisé conjointement car, en pratique, la réduction de dimension opère également une sorte de lissage ou filtre passe-bas. En effet, il est courant d'observer sur les derniers vecteurs propres les composantes les plus perturbées de la fonction aléatoire. Cela s'explique simplement car dans le cas d'un processus stationnaire ou « peu » éloigné de la stationnarité, son opérateur de covariance commute avec l'opérateur retard et possède donc les mêmes fonctions propres périodiques. L'ACP ressemble alors fortement à une décomposition en séries de Fourier et c'est pourquoi, dans les premiers travaux sur ce type de données, Deville [DEV 74] associait déjà ACP et analyse harmonique.

Les deux paramètres : dimension et lissage, interfèrent donc l'un sur l'autre. Plus précisément, la réduction de dimension permet de moins lisser à l'aide des splines et donc de trouver une valeur optimale de $\ell$ plus petite que celle qui serait obtenue avec le lissage seul. C'est une des raisons qui fait que cette ACP fonctionnelle conduit à de meilleures estimations des courbes qu'une succession de régression non paramétrique pour laquelle chaque paramètre de lissage serait optimisé indépendamment par validation croisée.

Un même critère, aidant au choix de dimension et proposé par Besse [BES 92] peut être utilisé. Il est basé sur une approximation du risque moyen quadratique mesurant la qualité d'estimation du sous-espace de représentation $E_q$ :

$$R_q = \mathbb{E}\frac{1}{2}\left\|\mathbf{P}_q - \widehat{\mathbf{P}_q}\right\|^2 = q - \mathrm{tr}\mathbf{P}_q\widehat{\mathbf{P}_q}$$

L'approximation par la théorie des perturbations de l'estimation jackknife est donnée par :

$$\widehat{R_{Pq}} = \frac{1}{n-1} \sum_{k=1}^{q} \sum_{j=k+1}^{p} \frac{\frac{1}{n}\sum_{i=1}^{n} c_{ik}^2 c_{ij}^2}{(\lambda_j - \lambda_k)^2} \tag{6.13}$$

où $c_{ij}$ désigne le terme général de la matrice $\mathbf{X}\mathbf{A}_\ell^{1/2}\mathbf{V}_q$.

### 6.3.3. *ACP de séries climatiques*

Besse et coll. [BES 97] ont montré sur des données simulées l'efficacité de cette approche associant dans le même problème d'optimisation des contraintes de réduction de rang et de régularité. Un lissage de chaque trajectoire prise séparément à base de validation croisée conduit à des résultats moins performants. La prise en compte de la structure de covariance à travers l'ACP permet une meilleure extraction du signal pour différents rapports signal sur bruit. On note encore que, lorsque la variance du bruit devient relativement importante c'est-à-dire plus grande que la ou les dernières valeurs propres de la partie signal, il est préférable de réduire la dimension en conséquence. Cette étude montre également que le critère $\widehat{R_{Pq}}$ de choix de dimension fournit des résultats suffisamment précis pour être opérationnels.

Nous préférons illustrer cette section par un exemple de données réelles particulièrement bruitées. Il s'agit des racines carrées des cumuls mensuels des précipitations de 26 villes en France observées pendant 10 ans [ECO 91]. La transformation (racine) s'avère nécessaire afin de stabiliser la variance comme dans le cadre d'un processus de Poisson. Pour traiter un problème suffisamment complexe on s'intéresse à 26×5 courbes observées durant 2 années consécutives. L'étude des données annuelles fournit le même type de résultats mais avec une composante en moins.

Une ACP classique calculée sur ces données fournit les résultats de la figure 6.2. Il s'agit donc des trois premiers vecteurs ou plutôt fonctions propres qui, très bruitées, sont difficiles à interpréter. Une ACP fonctionnelle incluant une contrainte de régularité contrôlé par le paramètre de lissage $\ell$ a ensuite été calculée. Le choix simultané de la dimension et de ce paramètre de lissage est guidé par les résultats de la figure 6.3. Celle-ci représente l'évolution de la stabilité du sous-espace de représentation en fonction de la valeur du paramètre de lissage et pour différentes dimensions. Cet indice $\widehat{R_{Pq}}$ lié au comportement d'écarts entre valeurs propres est très instable donc délicat à interpréter. Néanmoins, il apparaît que pour de petites valeurs de $\ell$ ($\log(\ell) < -5$), seule la première composante associée à une simple tendance est stable. Pour de plus grandes valeurs ($\log(\ell) > 6$), les données sont sur-lissées et beaucoup de composantes disparaissent. Le comportement de $R_{P5}$ présentant un minimum conduit finalement à retenir $q = 5$ et $\ell \approx 1$.

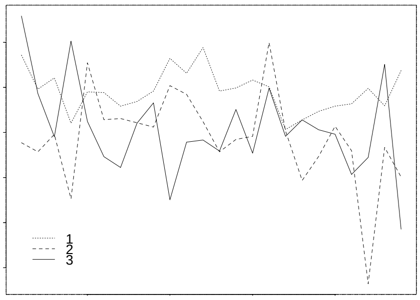


**Figure 6.2.** *Les trois premières fonctions propres de l'ACP classique (sans contrainte de régularité) des données pluviométriques. Très irrégulières, elles sont difficiles à interpréter*

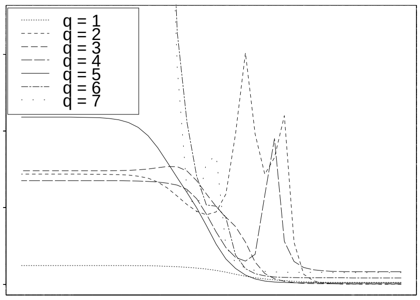


**Figure 6.3.** *Estimation de la stabilité* $\widehat{R_{Pq}}$ *du sous-espace de projection en fonction de* $\log(\rho)$ *et pour différents choix de dimension*

Les composantes principales d'une telle ACP avec contraintes de régularité devient alors nettement plus facile à interpréter à partir du graphique des fonctions propres plus régulières (figure 6.4) qui révèlent différentes composantes périodiques.

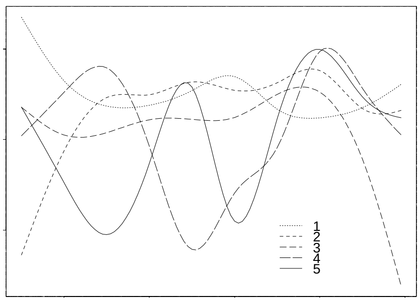


**Figure 6.4.** *Les cinq premières composantes principales engendrant le sous-espace de projection* $\widehat{\mathbf{P}}_q$

Un autre exemple d'ACP sur des séries climatiques, en l'occurrence des températures, est présenté par Antoniadou et coll. [ANT 01].

## 6.4. Modèle linéaire et extensions

La modélisation concerne l'estimation des paramètres fonctionnels de modèles statistiques pour lesquels la ou les variables explicatives sont des courbes. La première extension considérée dans la littérature a été de manière naturelle le modèle linéaire [RAM 91, HAS 93]). C'est par exemple le cas en chimiométrie ([FRA 93, GOU 98]) lorsqu'on souhaite expliquer et prédire la concentration de certaines molécules en fonction de réponses spectrométriques mesurées en plusieurs centaines de fréquences distinctes. Le nombre de variables explicatives (les fréquences de mesure) est alors généralement plus grand que la taille de l'échantillon et il devient nécessaire de réduire le nombre de variables explicatives, en particulier grâce à l'ACP fonctionnelle ou à la technique PLS étudiée dans le cadre fonctionnel par Preda et Saporta [PRE 02]. Une autre stratégie consiste à introduire une pénalisation dans l'estimation afin de stabiliser l'estimateur.

Ce problème de régression peut être modélisé de la façon suivante : soit $(Z, Y)$ un couple de variables aléatoires à valeurs dans $H \times \mathbb{R}$, $Z$ étant la variable explicative fonctionnelle (la courbe spectrométrique) et $Y$ la variable réponse réelle (la teneur en graisse). On peut supposer

$$\mathbb{E}[Y|Z = z] = \int_0^1 \psi(t)z(t)\,dt, \qquad z \in H, \tag{6.14}$$

où la fonction $\psi \in H$ est le cœfficient de régression fonctionnel à estimer.

Lorsque la variable à expliquer est discrète ou bornée, des extensions du modèle linéaire généralisé ont également été proposées par Eilers et Marx [MAR 96, MAR 99]) et Cardot *et coll.* [CAR 00b] Ces derniers prédisent l'occupation des sols à l'aide de mesures de l'évolution temporelle de données issues de télé-détection. Nous reviendrons sur cet exemple à la fin de la section.

Tous ces modèles s'étendent directement au cas de plusieurs variables explicatives fonctionnelles en considérant un nouvel espace fonctionnel produit d'espace fonctionnels.

### 6.4.1. *Modèle linéaire fonctionnel*

Notons $\Gamma$ l'opérateur de covariance de $Z$ et $\Delta$ l'opérateur de covariance croisée du couple $(Z, Y)$. C'est par définition l'opérateur :

$$\Delta x = \int_0^1 \mathbb{E}[Z(t)Y]\,x(t)dt, \quad x \in H.$$

En utilisant l'équation (6.14), il est facile de vérifier que

$$\Delta \;=\; \Psi\Gamma, \tag{6.15}$$

où $\Psi(x) =< \psi, x >$ .

Il faut donc inverser l'opérateur $\Gamma$ pour en déduire un estimateur de $\psi$. Le problème qui se pose alors (et se posera de manière identique pour l'estimation des processus auto-régressifs fonctionnels dans la section 6.5) provient du fait que l'inverse de $\Gamma$ n'est pas un opérateur borné. Pour contourner cette difficulté il devient nécessaire d'opérer une régularisation similaire à celles proposées dans le cadre des problèmes inverses. Nous présentons deux estimateurs différents qui tiennent compte de ce problème. Le premier, basé sur une réduction de la dimension, est un estimateur de la régression sur composantes principales lisses. Il consiste à inverser l'équation (6.15) dans l'espace engendré par les $q$ premiers vecteurs propres de l'opérateur de covariance empirique. Le second estimateur repose sur un critère des moindres carrés pénalisé.

Supposons maintenant que l'on dispose de l'échantillon $(\mathbf{z}_i, y_i),\ i = 1, \ldots, n$ où $\mathbf{z}_i = (z_i(t_j),\ j = 1, \ldots, p) \in \mathbb{R}^p$.

### 6.4.2. *Régression sur composantes principales (RCP)*

La régression sur composantes principales est une méthode classique en régression linéaire multiple lorsque les variables explicatives sont nombreuses et/ou corrélées. Elle consiste ici à estimer les fonctions propres de l'opérateur de covariance empirique (i.e. effectuer l'ACP fonctionnelle des courbes $z_i$) et à décomposer l'estimateur de la fonction $\psi$ sur cette base. De manière équivalente, cela revient à inverser l'équation (6.15) dans l'espace engendré par les fonctions propres associées aux plus grandes valeurs propres. Cette décomposition permet une réduction optimale de la dimension tout en conservant le maximum d'information, au sens de la variance, apportée par les variables explicatives. L'estimateur se décompose alors dans l'espace engendré par les fonctions propres estimées $\{\widehat{v}_j\}_{j=1,\ldots,q}$ :

$$\widehat{\boldsymbol{\psi}}(t) = \sum_{j=1}^{q} \beta_j \widehat{v}_j(t). \tag{6.16}$$

Le vecteur $\boldsymbol{\beta}$ des coefficients est déterminé en minimisant un critère des moindres carrés, les composantes principales $\langle \widehat{z}_i, \widehat{v}_j \rangle_{L^2} \;=\; \mathbf{z}_i' \mathbf{N} \mathbf{v}_j$ jouant le rôle de variables explicatives :

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^q} \sum_{i=1}^{n} w_i \left( Y_i - \sum_{j=1}^{q} \beta_j \mathbf{z}_i' \mathbf{N} \mathbf{v}_j \right)^2 .$$

Si on suppose que les trajectoires $z_i$ sont régulières, on peut anticiper la réduction de dimension en tenant compte de cette hypothèse de régularité par l'application de L'ACP fonctionnelle décrite dans la section précédente (eq. 6.10). Cette méthode d'approximation des fonctions propres de l'opérateur de covariance est reprise de manière plus détaillée dans la section 6.5. Il s'agit alors de construire un estimateur de l'opérateur d'auto-régression entre deux courbes successives.

### 6.4.3. *Approche par pénalisation*

Nous présentons ici une approche basée sur une régularisation de l'estimateur spline de régression [HAS 93, MAR 96, GOU 98]) comparable à la « ridge regression » en régression linéaire multiple. La régularisation ne porte plus sur la dimension $q$ comme dans la RCP mais directement sur l'estimateur de la fonction $\psi$. Nous supposons pour cela que c'est une fonction régulière et cherchons la solution des moindres carrés pénalisés

$$\min_{\boldsymbol{\psi}} \sum_{i=1}^{n} w_i \left(Y_i - \boldsymbol{\psi}'\mathbf{N}\mathbf{z}_i\right)^2 + \ell \, \|\psi\|_2^2 \,, \tag{6.17}$$

où $\boldsymbol{\psi} = (\psi(t_1), \ldots, \psi(t_p))'$. Le paramètre de lissage (ou régularisation) $\ell$ permet de contrôler la régularité de l'estimateur ainsi que sa stabilité. En termes matriciels, résoudre le problème (6.17) est équivalent à minimiser :

$$\min_{\boldsymbol{\psi}} \; \boldsymbol{\psi}'\mathbf{NSN}\boldsymbol{\psi} - 2\,\boldsymbol{\psi}'\mathbf{NZ}'\mathbf{DY} + \ell\,\boldsymbol{\psi}'\mathbf{M}\boldsymbol{\psi}. \tag{6.18}$$

La solution s'écrit alors

$$\widehat{\boldsymbol{\psi}} = \left(\mathbf{NSN} + \ell\mathbf{M}\right)^{-1} \mathbf{NZ}'\mathbf{DY}. \tag{6.19}$$

On reconnaît dans l'écriture de l'estimateur $\widehat{\boldsymbol{\psi}}$ l'inverse de l'opérateur de covariance pénalisé $\mathbf{NSN} + \ell\mathbf{M}$, ainsi que la transposée de l'opérateur de covariance croisée $\mathbf{NZ}'\mathbf{DY}$.

### 6.4.4. *Choix des paramètres de lissage*

Le choix d'un bon paramètre de lissage ou d'une bonne dimension est primordial mais difficile dans la pratique. Il contrôle en effet le compromis entre le biais de l'estimateur et sa variabilité. Un paramètre de lissage trop petit (ou une dimension trop grande pour la RCP) conduit à un estimateur qui s'adapte trop au données et dont le pouvoir prédictif est généralement mauvais. A l'opposé, si le paramètre de lissage est trop grand (ou la dimension de la RCP est trop petite) alors l'estimateur est stable mais biaisé et ne prend pas suffisamment en compte l'information apportée par la variable explicative. Des études asymptotiques [CAR 99a, CAR 99b]) ont montré que, pour obtenir des estimateurs convergents, le paramètre de lissage $\ell$ (resp. $q$) ne devait pas pas tendre trop rapidement vers zéro (resp. vers l'infini) lorsque la taille $n$ de l'échantillon tendait vers l'infini.

Dans la pratique, différentes stratégies issues des techniques de choix de modèles sont mises en œuvre. La validation croisée généralisée [WAH 90] initialement définie pour les splines de lissage peut être étendue aux estimateurs du modèle linéaire fonctionnel. Dans ce cadre, ce critère s'exprime :

$$\text{VCG} = \frac{\sum_{i=1}^{n}\left(y_i - \widehat{\boldsymbol{\psi}}'\mathbf{N}\mathbf{z}_i\right)^2}{\text{tr}\left(\mathbf{I}-\mathbf{A}\right)^2}$$

où la matrice $\mathbf{A}$ est la « hat matrix ». Elle dépend du paramètre de régularisation ($\ell$ ou $q$) et associe, par définition, au vecteur $\mathbf{y}$ le vecteur des réponses estimées $\widehat{\mathbf{y}}$. Pour l'estimateur par pénalisation spline, on a $\mathbf{A} = \mathbf{Z}\left(\mathbf{NSN}+\ell\mathbf{M}\right)^{-1}\mathbf{NZ}'\mathbf{D}$. Par ailleurs, on remarque que pour la RCP, la VCG s'exprime simplement puisque tr$(\mathbf{A}) = q$.

D'autres critères tels que la minimisation de l'erreur de prédiction sur échantillon test ou l'erreur quadratique pénalisée par le nombre de « variables » explicatives (Akaïke, BIC, ...) ont également été appliqués ([RAM 97, MAR 99]. Cardot *et al.* [CAR 02a] ont proposé un test de nullité du cœfficient fonctionnel basé sur le comportement asymptotique de l'opérateur de covariance croisée empirique. Ce test permet de tester la validité du modèle et l'égalité de la fonction $\psi$ à une fonction connue.

### 6.4.5. *Modèle linéaire généralisé et autres extensions*

Diverses extensions du modèle linéaire fonctionnel ont été proposées. La plus naturelle est de considérer un modèle linéaire généralisé afin d'expliquer des variables réponses $Y$ qualitatives, quantitatives discrètes ou bien encore bornées. Marx et Eilers [MAR 99] utilisent une approche par pénalisation similaire à celle étudiée dans la section 6.4.3 pour expliquer la probabilité de survie 2 ans après un cancer en fonction de la densité d'ADN dans les cellules cancéreuses. Le modèle s'écrit de manière formelle

$$I\!E[Y|Z=z] = g^{-1}\left(<\psi, z>\right), \qquad z \in H, \tag{6.20}$$

où $g$ est la fonction lien de type logit, log, ... Si $g$ est l'identité alors le modèle (6.20) est le modèle linéaire fonctionnel décrit plus haut.

L'estimateur de la fonction $\psi$ qui ne peut plus être obtenu de manière explicite est calculé en maximisant un critère de vraisemblance pénalisée

$$\max_{\boldsymbol{\psi}} \sum_{i=1}^{n} \log \mathcal{L}(y_i, \mathbf{z}_i, \boldsymbol{\psi}) - \ell\,\|\psi\|_2^2 \tag{6.21}$$

où $\mathcal{L}(.)$ est la vraisemblance. Marx et Eilers [MAR 99] utilisent une méthode de type moindres carrés pondérés itératifs pour trouver l'optimum [FAH 94] pour une description de cette méthode dans le cadre classique du modèle linéaire généralisé). La valeur du paramètre de lissage est ensuite choisie en minimisant un critère d'information de type BIC ou Akaïke. Certains résultats de convergence ont été obtenus par Cardot et Sarda [CAR 02b].

Une autre approche, similaire à la RCP, consiste à réduire la dimension en décomposant l'estimateur de $\psi$ sur les fonctions propres lisses de l'opérateur de covariance empirique des $z_i$. Les paramètres (*i.e* les coordonnées dans la base des fonctions propres) sont ensuite déterminés par maximum de vraisemblance. C'est ce point de vue que nous adoptons dans l'exemple de la section suivante.

Enfin, apparaissent maintenant des modèles semi et non paramétriques pour données fonctionnelles. Ferraty et Vieu [FER 02b] et Ferraty *et coll.* [FER 02a] montrent, sous certaines hypothèses, la convergence de l'estimateur à noyau dans le cadre de la régression avec une variable explicative fonctionnelle et Yao et Ferré [YAO 00] ont étendu l'approche « sliced inverse regression ».

#### 6.4.6. *Estimation du plan d'occupation des sols en fonction de l'évolution temporelle de mesures de télé-détection*

Cet exemple est tiré d'un travail effectué à l'INRA dans le cadre du programme UIC-Végétation. Pour plus de détails, on pourra se reporter à Cardot *et coll.* [CAR 00b].

Le capteur Végétation du satellite SPOT4 fournit de manière quasi quotidienne (haute répétitivité temporelle) des images de l'Europe (2000km $\times$ 2000km) à faible résolution spatiale (chaque pixel représente une zone de 1km $\times$ 1km). Les informations fournies par ce capteur dans le visible, le proche et le moyen infra-rouge, permettent de caractériser l'état de développement de la végétation à l'échelle d'une région. Les pixels observés sont *mixtes* ; ils mélangent différentes informations au sens où ils représentent différentes parcelles agricoles (dont la taille en France est nettement inférieure au km$^2$) et donc différentes cultures ou thèmes d'intérêt (blé, maïs, orge, forêt...).

Il s'agit d'utiliser les courbes d'évolution temporelle des pixels sur les images pour séparer les différents types de culture et déterminer le plan d'occupation des sols (POS) de la région, c'est-à-dire la proportion de chaque thème cultural à l'intérieur de chaque pixel de 1 km$^2$. Pour cela, nous disposons d'une petite région sur laquelle nous avons accès à un plan d'occupation des sols. Un enjeu de ce travail est que l'estimation de l'occupation des sols de régions avoisinantes permettrait d'en déduire une prédiction des productions sur une région plus vaste grâce à un couplage des données satellites et des données météorologiques avec un modèle de croissance.

Le problème statistique posé est l'estimation de proportions à partir d'observations longitudinales. Nous avons mis en œuvre une extension du modèle multilogit aux données fonctionnelles. L'intérêt d'une telle modélisation réside essentiellement dans la possibilité de considérer des courbes de réflectance dont les instants de mesure varient d'un pixel à l'autre, ce qui est le cas notamment lorsqu'une partie de l'image est cachée par les nuages.

Nous noterons $\boldsymbol{\pi}_i = [\pi_{ik}]_{k=1,\ldots,g}$ l'occupation des sols (POS) du pixel $i$ : c'est le vecteur des proportions de chaque culture $k,\ k = 1,\ldots,g$, à l'intérieur du pixel $i$. Par construction $\pi_{ik} \geq 0$ et $\sum_k \pi_{ik} = 1$ pour $i = 1,\ldots,n = 1554$. Sur la zone d'intérêt, d'environ 40km × 40km, on dispose également des images à haute fréquence temporelle fournies par le capteur Végétation de Spot dans 4 longueurs d'onde, Bleu, Rouge (R), Proche Infra-Rouge (PIR) et Moyen Infra-Rouge (MIR), sur une période allant de mars 1998 à août 1998. Deux indices couramment utilisés par les bioclimatologues ont également été construits, $NDVI = (PIR-R)/(PIR+R)$ et $PVI = (PIR - 1.2R)/\sqrt{2.2}$.

Les courbes d'évolution discrètes de la réflectance, observées aux instants $t_1$, …, $t_k$, …, $t_p$, sont notées $\mathbf{z}_i = [z_i(t_1),\ldots,z_i(t_p)]^T$ , pour chaque pixel $i$. Après suppression des dates où la couverture nuageuse était trop importante, on dispose de 32 dates exploitables ($p = 32$).

On suppose que la proportion $\pi_{ik}$ de la culture $k$ sur le pixel $i$ est issue d'une loi multinomiale dont les paramètres vérifient :

$$\mathbb{E}(\pi_{ik}|z_i) = \frac{\exp\left(\delta_k + \int_T \psi_k(t)\,(z_i(t) - \mathbb{E}(z_i(t)))\ dt\right)}{\sum_{l=1}^{p} \exp\left(\delta_l + \int_T \psi_l(t)\,(z_i(t) - \mathbb{E}(z_i(t)))\ dt\right)}. \tag{6.22}$$

Pour des raisons d'identifiabilité, on pose $\psi_g = 0$ et $\delta_g = 0$. Le régresseur $\psi_k$ s'interprète par comparaison avec la classe de référence $g$.

Le nombre de paramètres à estimer, qui dépend du nombre de classes du POS et du nombre de variables explicatives, est très important ($32 \times (10-1)$). Le grand nombre de variables explicatives (le nombre de dates, entre 30 et 40 pour notre exemple) nous conduit à réduire la dimension.

Une première étape dans la phase d'estimation consiste donc à effectuer l'analyse en composantes principales *fonctionnelle* des courbes discrétisées $\mathbf{z}_i, i = 1,\ldots,n$, afin de réduire la dimension des variables explicatives de manière optimale au sens de la variance. On obtient ainsi une décomposition des fonctions de régression $\psi_k$

similaire à (6.16). Ensuite, les cœfficients associés aux estimateurs des régresseurs sont estimés par maximum de vraisemblance. Nous avons choisi la classe « urbain » comme référence en raison de sa stabilité au cours du temps.

Une deuxième étape consiste ensuite à sélectionner les composantes principales « significatives » à l'aide du test du rapport de vraisemblance. Celui-ci est construit de manière ascendante sur les composantes principales.

Les 1554 pixels observés ont été partagés en deux groupes. Un échantillon d'apprentissage constitué de 1055 pixels obtenus par sondage simple sans remise. Il est utilisé pour calculer les estimateurs. Un échantillon test constitué des 499 pixels restants sur lesquels on détermine le « meilleur » modèle et pour lequel il est possible d'évaluer l'erreur de prédiction. Le modèle $M_0$ qui prédit l'occupation des sols par la moyenne empirique sert de référence.

Les estimations des fonctions de régression associées au canal NDVI sont tracées dans la Figure 6.5. On note la grande ressemblance, à une constante près, des courbes

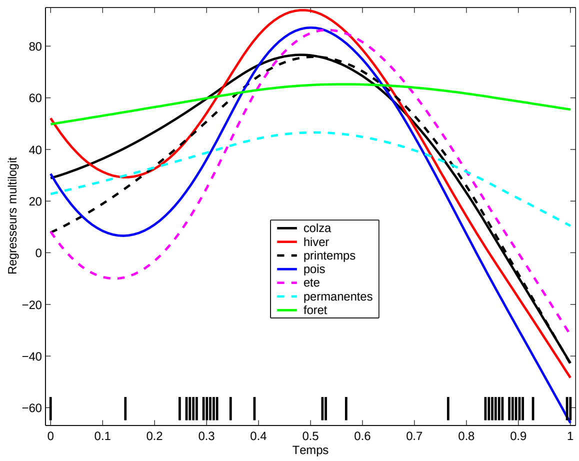


**Figure 6.5.** *Fonctions de régression associées au canal NDVI pour le modèle multilogit fonctionnel*

associées aux thèmes « permanentes » et « forêt ». On constate aussi que les erreurs de prédiction selon le critère

$$R_{ik} = \frac{|\pi_{ik} - \widehat{\pi_{ik}}|}{\frac{1}{n}\sum_{i=1}^{n} \pi_{ik}}$$

des modèles multilogit sont systématiquement plus faibles que celles du modèle $M_0$ (Table 6.1). Ainsi, l'évolution temporelle des images apporte une information sur le contenu des pixels. Le canal PVI semble le plus efficace pour prévoir le POS, il donne les meilleurs résultats pour cinq thèmes sur dix.

| Classes | NDVI | PVI | Bleu | Rouge | PIR | MIR | $M_0$ |
|---|---|---|---|---|---|---|---|
| urbain | 0.49 | **0.36** | 0.47 | 0.54 | 0.41 | 0.51 | 0.86 |
| eau | 0.43 | **0.29** | 0.78 | 0.62 | 0.61 | 0.31 | 1.30 |
| colza | 0.48 | 0.46 | **0.45** | 0.50 | 0.47 | 0.47 | 0.59 |
| hiver | 0.20 | 0.21 | **0.19** | 0.20 | 0.22 | **0.19** | 0.30 |
| printemps | 0.58 | **0.56** | 0.60 | 0.61 | 0.65 | 0.61 | 0.69 |
| pois | 0.50 | **0.43** | 0.45 | **0.43** | 0.48 | 0.46 | 0.63 |
| été | 0.61 | 0.68 | 0.61 | 0.60 | 0.76 | **0.53** | 0.88 |
| permanentes | 0.47 | **0.46** | 0.52 | 0.49 | **0.46** | 0.50 | 0.61 |
| forêt | 0.34 | 0.36 | 0.34 | **0.31** | 0.45 | 0.35 | 0.98 |
| pomme de terre | 0.90 | 0.93 | 0.94 | 0.90 | 1.06 | **0.85** | 1.31 |

**Tableau 6.1.** *Risques médians sur l'échantillon test pour l'approche basée sur le modèle Multi-logit et le modèle de référence $M_0$. Les chiffres en gras correspondent aux meilleurs prédicteurs par classe d'occupation*

## 6.5. Prévision

Cette section s'intéresse à la prévision d'un processus présentant naturellement une composante périodique. C'est le cas des données étudiées dans la section 6.5.6 et issues de l'observation mensuelle du phénomène *El Niño* durant plusieurs années. Plutôt que de prévoir séquentiellement une série de valeurs comme le ferait une modélisation SARIMA, la méthode développée propose une *prévision globale* sur la période considérée c'est-à-dire une année. Les modèles et exemples présentés sont issus essentiellement des articles de Bosq [BOS 91], Besse et Cardot [BES 96] et Besse et coll. [BES 00]. Ces modèles ont également été appliqués avec un certain succès à la prévision de la pollution par l'Ozone [DAM 02].

L'objectif est donc la prévision d'un processus $(X_t)_{t\in\mathbb{R}}$ à temps continu, pas nécessairement stationnaire, sur un intervalle de temps $\delta$, dont on fixe la valeur $\delta = 1$ sans perte de généralité. Cet intervalle correspond à la période principale du processus. On définit alors, à partir de $(X_t)$, un processus à temps discret $(Z_n)_{n\in\mathbb{Z}}$, supposé du second ordre, prenant ses valeurs dans un espace fonctionnel, par exemple $L^2[0,1]$ l'espace des fonctions de carré intégrable sur $[0,1]$, en découpant les trajectoires sur des tronçons de longueur $\delta = 1$.

### 6.5.1. *Processus Auto-régressifs Hilbertiens d'ordre 1*

Le cadre théorique de cette approche a été proposé par Bosq [BOS 91] puis complété et synthétisé par Bosq [BOS 00]). Il s'intéresse à des processus fonctionnels $(Z_n)$ auto-régressifs d'ordre 1. Ces processus, lorsqu'ils sont à valeurs dans un espace de Hilbert comme par exemple $L^2[0,1]$, sont notés ARH(1). Le processus $(Z_i)_{i\in\mathbb{Z}}$ est un processus auto-régressif hilbertien du premier ordre d'espérance $a \in H$ et d'opérateur d'auto-corrélation $\rho$ s'il vérifie l'équation :

$$\forall i \in \mathbb{Z}, \quad Z_i - a = \rho(Z_{i-1} - a) + \epsilon_i. \tag{6.23}$$

L'opérateur d'auto-corrélation $\rho$ est supposé compact et vérifie $\sum_{n\geq 1}\|\rho^n\| < +\infty$. Le terme d'innovation $\{\epsilon_i\}$ est centré i.i.d. dans $H$ de variance finie, $\mathbb{E}\|\epsilon_i\|_H^2 = \sigma^2 < +\infty$.

Dans ce cas, l'estimation de l'espérance conditionnelle $\mathbb{E}(Z_{i+1}|Z_i, Z_{i-1}, \ldots)$ nécessaire à la prévision se ramène à celle de l'opérateur $\rho$ :

$$\mathbb{E}(Z_{i+1}|Z_i, Z_{i-1}, \ldots) - a = \rho(Z_i - a)\ , \ \ i \in \mathbb{Z}. \tag{6.24}$$

Dans l'espace $L^2[0,1]$, cet opérateur admet une écriture sous forme d'opérateur intégral de noyau $\rho(s,t)$ et l'équation (6.24) s'exprime de manière équivalente :

$$\mathbb{E}(Z_{i+1}(t)|Z_i, Z_{i-1}, \ldots) - a(t) = \int_0^1 \rho(s,t)(Z_i - a)(s)\ ds, \quad t \in [0,1],\ i \in \mathbb{Z}. \tag{6.25}$$

Les opérateurs de covariance $\Gamma = \mathbb{E}((Y_i - a)\otimes(Y_i - a))$ $\Delta = \mathbb{E}((Y_i - a)\otimes(Y_{i+1} - a))$ de ce processus vérifient une équation similaire à l'équation (6.15) obtenue pour le modèle linéaire fonctionnel :

$$\Delta = \rho\Gamma. \tag{6.26}$$

L'opérateur $\Delta$ admet également une écriture sous forme d'opérateur intégral :

$$\Delta f(t) = \int_0^1 \gamma(s, t+1) f(s)\ ds, \quad t \in [0,1],\ f \in L^2[0,1]. \tag{6.27}$$

Comme pour le modèle linéaire fonctionnel, il s'agit d'inverser l'opérateur de covariance $\Gamma$ pour construire un estimateur de l'opérateur $\rho$. L'inverse de $\Gamma$ n'étant pas un opérateur borné, Bosq [BOS 91] a proposé la démarche suivante afin de construire un estimateur de $\rho$ :

1) analyse spectrale de l'opérateur $\Gamma$ fournie par les solutions de l'équation intégrale

$$\int_{[0,1]} \gamma(s,t) f(s) ds = \lambda f(t)\ ,\ t \in [0,1]\ \ ;$$

2) projection des trajectoires du processus sur le sous-espace $H_q$ de dimension finie $q$ engendré par les fonctions propres $\phi_1, \ldots, \phi_q$ de $\Gamma$ associées aux $q$ plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_q > 0$ ;

3) inversion de la restriction de $\Gamma$ à $H_q$ ;

4) approximation de rang fini d'un estimateur de $\rho$ à partir de (6.26).

De manière similaire à la régression sur composantes principales (6.16), l'estimateur de $\widehat{\rho}$ est approché dans la base engendrée par les premières fonctions propres de l'opérateur de covariance :

$$\widehat{\rho}(s,t) \sum_{j=1}^{q} \sum_{j'=1}^{q} \widehat{\beta}_{jj'} \widehat{v}_j(t) \widehat{v}_{j'}(s). \tag{6.28}$$

Dans la pratique, les trajectoires sont observées en des points de discrétisation et il se pose alors le problème de l'approximation nécessaire des estimateurs et de la prévision. La démarche proposée par Bosq [BOS 91] et Pumo [PUM 92] repose sur deux points importants :

– une estimation de l'opérateur $\rho$ de rang fini ;

– une approximation des trajectoires du processus par interpolation linéaire entre les points de discrétisation.

Différentes stratégies peuvent être envisagées pour approcher les vecteurs propres de l'opérateur de covariance. Nous proposons de remplacer l'interpolation linéaire par une estimation non paramétrique des trajectoires avant d'entamer l'approximation de l'opérateur de covariance $\Gamma$. De plus, sachant que cette approximation doit être de rang fini pour être inversée, on propose d'anticiper cette réduction de rang en recherchant des approximations des trajectoires dans un sous-espace fonctionnel de dimension finie $q$. Cette approche nécessite donc de définir une estimation non paramétrique des trajectoires sous contrainte de rang. Le lissage spline qui est défini par un problème d'optimisation semble, dans ce cas, préférable à une méthode de convolution (noyau). La solution de ce problème présentée dans la section 3 dans le cadre de l'estimation fonctionnelle simultanée de plusieurs observations d'un processus, est donc intégrée dans le modèle de prévision.

### 6.5.2. *ARH(1) lisse*

On suggère donc d'anticiper la contrainte de rang en estimant simultanément les trajectoires supposées lisses du processus dans un espace de dimension réduite $q$. Cette approche qui a été proposée par Besse et Cardot [BES 96] s'est révélée plus performante sur des simulations ainsi que sur un exemple de prévision du trafic auto-routier que l'interpolation linéaire [PUM 92] et que le lissage des trajectoires suivi de l'estimation de l'opérateur de covariance. On considère pour cela le problème d'optimisation suivant dont la solution est fournie dans la proposition 6.3.1 :

$$\min_{\hat{z}_i \in H_q} \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{p} \sum_{j=1}^{p} \left(z_i(t_j) - \widehat{z}_i(t_j)\right)^2 + \ell \left\| D^2 \widehat{z}_i \right\|_{L^2}^2 \right) \tag{6.29}$$

avec $H_q$ sous-espace de $H$ de dimension $q$. L'espérance du processus $Z_i$ est estimée par interpolation spline aux instants de mesure $(t_1, \ldots, t_j)$ du vecteur $\widehat{\mathbf{a}} = \mathbf{A}_\ell \bar{\mathbf{z}}$. En notant $\widehat{\mathbf{z}}_i = \widehat{\mathbf{a}} + \widehat{\mathbf{x}}_i$ la solution du problème (6.29), on en déduit la représentation matricielle des estimateurs des opérateurs de covariance :

$$\widehat{\mathbf{\Gamma}}_{q,\ell} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{x}}_i \widehat{\mathbf{x}}_i' \mathbf{N}, \tag{6.30}$$

et

$$\widehat{\boldsymbol{\Delta}}_{q,\ell} = \frac{1}{n-1} \sum_{i=1}^{n-1} \widehat{\mathbf{x}}_{i+1} \widehat{\mathbf{x}}_i' \mathbf{N}. \tag{6.31}$$

Il suffit ensuite de construire l'inverse généralisé, noté $\widehat{\boldsymbol{\Gamma}}_{q,\ell}^{-1}$, de la matrice $\widehat{\boldsymbol{\Gamma}}_{q,\ell}$ pour en déduire un estimateur de l'opérateur $\rho$ sous forme matricielle :

$$\widehat{\boldsymbol{\rho}}_{q,\ell} = \widehat{\boldsymbol{\Gamma}}_{q,\ell}^{-1} \, \widehat{\boldsymbol{\Delta}}_{q,\ell}. \tag{6.32}$$

Finalement, la prévision de $\mathbf{z}_{n+1}$ est donnée par la formule :

$$\widehat{\mathbf{z}}_{n+1} = \widehat{\boldsymbol{\rho}}_{q,\ell} \widehat{\mathbf{x}}_n + \mathbf{A}_\ell \bar{\mathbf{z}}. \tag{6.33}$$

Dans la pratique le choix des valeurs des paramètres de « régularisation » $\ell$ et $q$ est très important pour obtenir de bonnes prévisions. Le choix de ces valeurs est discuté dans la section 6.5.4.

Notons enfin que si les instants de discrétisation sont différents d'une courbe à l'autre il est possible d'adapter l'estimateur proposé en (6.29) en approchant au préalable les trajectoires sur une base de fonctions (les B-splines par exemple) puis en considérant un problème d'optimisation similaire à partir des coordonnées dans cette base de fonctions [BES 97].

### 6.5.3. *ARH(1) Local*

Dans bien des situations l'hypothèse de stationnarité du processus fonctionnel $(Z_i)_{i\in\mathbb{Z}}$ est une hypothèse trop forte. C'est par exemple le cas pour la série El Niño où on observe certains évènements exceptionnels. Besse et coll. [BES 00] ont proposé des approximations locales des opérateurs de covariance permettant de s'affranchir de cette hypothèse. Cette approche revient à effectuer une approximation linéaire locale d'un modèle auto-régressif Hilbertien non linéaire. Ils définissent pour cela des estimateurs à noyau des opérateurs de covariance permettant de tenir compte de la proximité de l'année en cours avec les années passées dans la procédure d'estimation. Considérons un noyau réel gaussien $K_h$ :

$$K_h(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2h}\right) \tag{6.34}$$

et définissons les estimateurs locaux des opérateurs de covariance :

$$\widehat{\Delta}_h(\mathbf{x}) = \frac{\sum_{i=1}^{n-1} K_h\left(\|\mathbf{x}_i - \mathbf{x}\|_{\mathbf{N}}\right) \mathbf{x}_i \mathbf{x}_{i+1}' \mathbf{N}}{\sum_{i=1}^{n-1} K\left(\|\mathbf{x}_i - \mathbf{x}\|_{\mathbf{N}}\right)}, \tag{6.35}$$

$$\widehat{\Gamma}_h(\mathbf{x}) = \frac{\sum_{i=1}^{n} K_h\left(\|\mathbf{x}_i - \mathbf{x}\|_{\mathbf{N}}\right) \mathbf{x}_i \mathbf{x}_i' \mathbf{N}}{\sum_{i=1}^{n-1} K_h\left(\|\mathbf{x}_i - \mathbf{x}\|_{\mathbf{N}}\right)}. \tag{6.36}$$

Définissons la matrice de poids $\mathbf{W}_h(\mathbf{x})$ d'éléments diagonaux :

$$w_{i,h}(\mathbf{x}) = \frac{K_h\left((\mathbf{x}_i - \mathbf{x})'\mathbf{N}(\mathbf{x}_i - \mathbf{x})\right)}{\sum_{l=1}^{n} K\left((\mathbf{x}_l - \mathbf{x})'\mathbf{N}(\mathbf{x}_l - \mathbf{x})\right)}, \quad i = 1, \ldots, n. \tag{6.37}$$

Alors, les opérateurs de covariance s'expriment simplement

$$\widehat{\Delta}_h(\mathbf{x}) = \sum_{i=1}^{n} w_{i,h}(\mathbf{x})\ \mathbf{x}_i \mathbf{x}_{i+1}' \mathbf{N} \quad \text{et} \quad \widehat{\Gamma}_h(\mathbf{x}) = \mathbf{X}'\mathbf{W}_h(\mathbf{x})\mathbf{X}\mathbf{N}.$$

L'estimateur de $\rho$ est alors défini en effectuant l'analyse spectrale de la matrice $\widehat{\Gamma}_h$ évaluée au point $\mathbf{x} = \mathbf{z}_n - \widehat{\mathbf{a}}$. L'inverse de $\widehat{\boldsymbol{\Gamma}}_h$ de rang $q$ s'écrit

$$\left(\widehat{\mathbf{P}}_q \widehat{\boldsymbol{\Gamma}}_h \widehat{\mathbf{P}}_q\right)^{-1} = \sum_{l=1}^{q} \frac{1}{\hat{\lambda}_l} \widehat{\mathbf{v}}_l \widehat{\mathbf{v}}_l' \mathbf{N}$$

où $(\hat{\mathbf{v}}_l, l = 1, \ldots, q)$ sont les vecteurs propres de $\widehat{\boldsymbol{\Gamma}}_h$ pour la métrique $\mathbf{N}$ associés aux plus grandes valeurs propres $(\hat{\lambda}_l, l = 1, \ldots, q)$. En inversant l'équation liant les opérateurs $\rho$, $\Delta$ et $\Gamma$ au point $\mathbf{z}_n - \widehat{\mathbf{a}}$ on définit ensuite un estimateur local de l'opérateur de corrélation $\rho$ par :

$$\widehat{\boldsymbol{\rho}}_{q,h} = \left(\sum_{l=1}^{q} \hat{\mathbf{v}}_l \hat{\mathbf{v}}_l'\right) \mathbf{N}\widehat{\boldsymbol{\Delta}}_h \left(\sum_{l=1}^{q} \frac{1}{\hat{\lambda}_l}\, \hat{\mathbf{v}}_l \hat{\mathbf{v}}_l'\right) \mathbf{N}. \tag{6.38}$$

Cet estimateur dépend des paramètres de lissage $q$ et $h$ qui sont déterminés en minimisant l'erreur de prévision sur un échantillon test (cf. section 6.5.4). Pour effectuer une prévision de $\mathbf{z}_{n+1}$, il suffit ensuite d'appliquer la formule (6.33).

Une variante de cette méthode consiste à effectuer au préalable un lissage spline des trajectoires plutôt qu'une interpolation lorsque celles-ci sont bruitées. Elle présente l'inconvénient de dépendre de trois paramètres de « lissage » dont il faut ensuite déterminer les valeurs.

### 6.5.4. *Choix des paramètres de lissage*

Chaque procédure décrite précédemment fait intervenir un ou plusieurs paramètres de lissage ou régularisation dont la valeur a une forte incidence sur la qualité de la prévision. Ainsi, pour le prédicteur ARH(1) lisse, un trop grand paramètre de lissage $\ell$ entraîne un sur-lissage des trajectoires et par conséquent un mauvais estimateur de l'opérateur $\rho$.

Il est donc important de proposer une méthode automatique (ou presque) de choix de ces paramètres. Les techniques de choix de modèles utilisées dans le cadre du modèle linéaire fonctionnel ne sont plus valides en raison de la dépendance entre les trajectoires $Z_i$ du processus. Ces paramètres sont donc choisis dans ce cadre en séparant l'échantillon en deux sous-échantillons, un échantillon d'apprentissage constitué des premiers tronçons de trajectoires et un échantillon test. Les estimateurs sont construits sur l'échantillon d'apprentissage puis évalués et comparés à l'aide d'un critère mesurant l'erreur quadratique de prévision sur l'échantillon test. Les paramètres de lissage sont choisis de manière à donner les estimateurs qui fournissent les meilleures prévisions sur l'échantillon test.

### 6.5.5. *Résultats asymptotiques*

Un certain nombre de résultats de convergence pour les estimateurs des processus ARH(1) ont été démontré depuis le début des années 90. En l'absence de discrétisation des trajectoires observées, Bosq [BOS 91] a prouvé la convergence de l'estimateur de l'opérateur de covariance à la vitesse $1/n$ et à montré que l'on obtenait un estimateur convergent de $\rho$ sous certaines hypothèses et en particulier sous la condition que la dimension $q$ tende suffisamment lentement vers l'infini avec taille de l'échantillon. Par exemple, si le bruit $\epsilon_i(t)$ est un mouvement brownien alors $q$ doit être de l'ordre de $\log n$. Ensuite Mourid [MOU 95] a défini les processus AR Hilbertiens d'ordre supérieur et montré la convergence des estimateurs de l'opérateur d'auto-corrélation en utilisant des techniques similaires à celles de Bosq (1991). Lorsque les trajectoires sont discrètes, Pumo [PUM 92] a prouvé la convergence de la méthode basée sur l'interpolation linéaire lorsque le pas de discrétisation tend suffisamment vite vers zéro. Enfin Cardot [CAR 98] a prouvé sous des hypothèses similaires sur le schéma de discrétisation la convergence des estimateurs dans les modèles ARH(p) basés sur l'approximation spline.

Actuellement les recherches s'orientent vers la construction de tests d'adéquations et l'obtention de convergence en lois asymptotiques afin de pouvoir associer un intervalle de confiance aux prévisions [BOS 00, MAS 00]).

### 6.5.6. *Prévisions de séries climatiques*

Besse et Cardot [BES 96] ont montré l'intérêt d'une approche fonctionnelle de la prévision sur des données simulées et des données de trafic auto-routier. Comme pour l'ACP fonctionnelle, la double considération d'un lissage et d'une réduction de la dimension aux $q$ premières composantes principales contribue à une amélioration sensible des résultats et ainsi des erreurs de prévisions. Cette section reprend brièvement un exemple de Besse et col. [BES 00] pour illustrer et comparer les modèles proposés avec d'autres plus classiques.

ENSO (El Niño Southern Oscillation) est un phénomène climatique majeur d'interaction entre l'atmosphère et l'océan Pacifique dans la zone tropicale. El Niño (EN) se caractérise par des variations importantes de la température, particulièrement en décembre, très corrélées à des oscillations de la pression atmosphérique (SO). EN est donc observé à travers l'évolution de températures moyennes mensuelles mesurées au large du Chili à la surface de l'océan tandis que SO est mesurée par la pression atmosphérique au niveau de la mer à Tahiti. Ces séries sont connues depuis 1950 et considérées dans cette étude jusqu'en 1996. Elles sont représentées dans la figure 6.6.

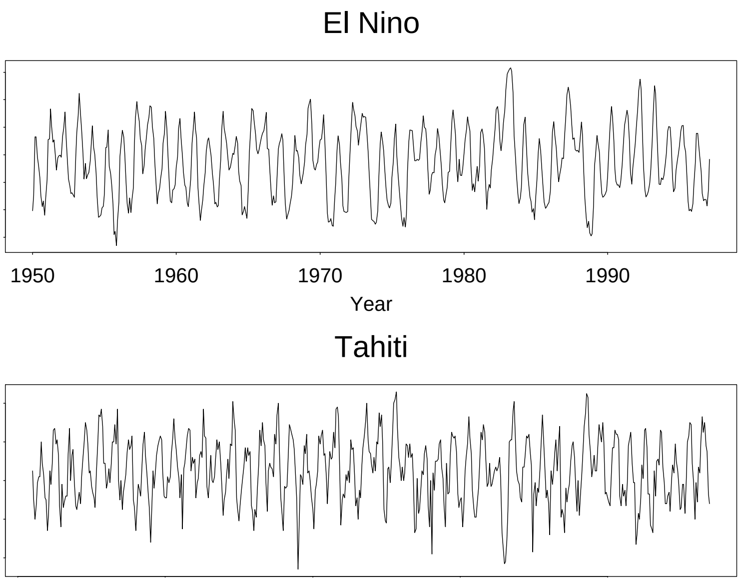


**Figure 6.6.** *Séries des moyennes mensuelles de l'index de température (Nino-3) à la surface de la mer et des moyennes mensuelles de la pression atmosphérique à Tahiti*

Différents prédicteurs sont comparés sur ces données en utilisant le même protocole de validation croisée pour l'optimisation des paramètres spécifiques. Ainsi, les 36 premières années servent d'apprentissage, les paramètres sont optimisés pour prévoir les années 32 à 36. Les méthodes sont comparées entre elles par leur prévision des années 37 à 46 et avec une prévision élémentaire qui est la simple moyenne climatique.

Les meilleurs modèles paramétriques obtenus de la sorte ont été un SARIMA $(0,1,1) \times (1,0,1)_{12}$ pour les températures (EN), et un SARIMA $(1,1,1) \times (0,1,1)_{12}$ pour les pressions (SO). Tous deux ont satisfait au test du porte manteau. Les autres méthodes de prévision considérées sont non paramétriques :

**Noyau** ou prévision non paramétrique par la méthode du noyau usuelle répétée pour des horizons de 1 à 12 mois. Les largeurs de fenêtre optimales sont respectivement 0.9 et 1.5 pour EN et SO.

**Noyau fonct.** est également une prévision par régression non paramétrique mais vectorielle pour prévoir les 12 valeurs simultanément. Valeurs optimales respectivement 0.3 et 0.6.

**ARH(1) lisse** requiert 4 (resp. 3) dimensions associées à un paramètre de lissage de $1.610^{-5}$ (resp. $810^{-5}$).

**ARH(1) local** requiert 3 (resp. 2) dimensions associées à un paramètre de lissage de $0.910^{-5}$ (resp. $510^{-5}$).

La série des pressions est sensiblement plus irrégulières, ce qui n'a rien de surprenant, et nécessite donc des lissages plus importants que la série des températures. La figure 6.7 montre une comparaison graphique de la prévision de l'année 1986 de El Niño. Les prévisions scalaires y sont reliées par des segments de droite tandis que les

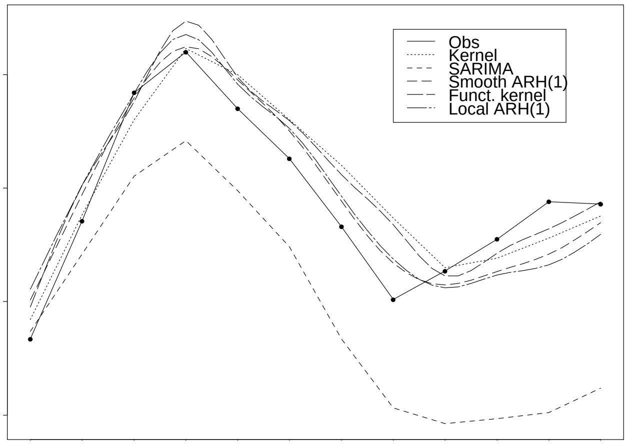


**Figure 6.7.** *Comparaison de l'année 1986 de El Niño avec différentes prévisions*

prévisions fonctionnelles sont interpolées par des splines cubiques. Pour cette année, le modèle SARIMA se montre très défaillant tandis que les prédicteurs non paramétriques ont des performances voisines. des résultats similaires sont obtenus pour la série SO. Le tableau 6.2 synthétise ces performances sur une période de 10 ans en prenant en compte l'erreur quadratique moyenne de prévision (EQMP) et l'erreur absolue moyenne relative (EAMR).

| | El Niño index | | S. Osc. index | |
|---|---|---|---|---|
| Prédicteur | EQMP | EAMR | EQMP | EAMR |
| Climat. | 0.73 | 2.5 % | 0.91 | 6.3 % |
| SARIMA | 1.45 | 3.7 % | 0.95 | 6.2 % |
| Noyau | 0.60 | 2.3% | 0.87 | 6.1% |
| Noyau fonct. | 0.58 | 2.2% | 0.82 | 6.0% |
| ARH(1) lisse | 0.55 | 2.3% | **0.78** | **5.8%** |
| ARH(1) local | **0.53** | **2.2%** | 0.82 | 5.8% |

**Tableau 6.2.** *Comparaisons des erreurs quadratiques et erreurs absolues relatives de prévision sur la période 1987-1996*

Dans l'ensemble, les méthodes non paramétriques se révèlent bien meilleures que la méthode paramétrique (SARIMA) qui est même inférieure à la prévision climatique. Plus précisément, les méthodes fonctionnelles apparaissent sur cet exemple sensiblement plus performantes que les approches par noyau scalaire ou vectorielle. Le manque de stationnarité est sans doute responsable du mauvais comportement du modèle SARIMA surtout dans le cas de la série des températures (EN) dont la meilleure prévision est fournie par le modèle ARH(1) avec une estimation locale de la structure de covariance.

Il n'y a pas lieu de généraliser ces résultats ; néanmoins ils montrent que les modèles fonctionnels de prévision présentés peuvent se montrer efficaces et sont donc à intégrer dans la boîte à outil standard du statisticien en concurrence avec d'autres plus classiques.

## 6.6. Conclusion

Tant les exemples traités dans ce chapitre que l'analyse de données simulées montrent l'intérêt de prendre en compte le caractère fonctionnel des données observées. Cette approche implique un coût méthodologique. C'est l'effort qui permet d'adapter au mieux les méthodes multidimensionnelles aux données et à leur problématique afin d'en extraire la part la plus pertinente. Toute méthode statistique exploratoire, de classification ou de modélisation, a été ou peut être adaptée à ce contexte en intervenant à quatre niveaux.

1) Plongement des observations dans un espace fonctionnel par l'utilisation d'une base idoine : splines, Fourier, ondelettes. S'arrêter à un simple changement de base dans $\mathbb{R}^p$ ne suffit pas à réellement prendre en compte le caractère fonctionnel qui suppose une certaine régularité des trajectoires observées éventuellement avec du bruit.

2) Nécessité éventuelle de recaler les courbes par une transformation non linéaire de l'échelle des temps (*curve registration*) afin de faire coïncider dans le temps certains évènements.

3) Intervention d'un lissage susceptible de réduire le bruit et/ou d'une norme appropriée (Sobolev) prenant en compte les hypothèses de régularité ou une connaissance *a priori* du phénomène. Cette étape permet également de stabiliser l'estimateur

4) Dans la mesure du possible, optimisation conjointe plutôt que séquentielle des critères et paramètres impliqués dans l'analyse.

**Remerciements**. Nous souhaitons remercier tous les participants au groupe de travail STAPH de Toulouse pour les discussions enrichissantes autour de la statistique fonctionnelle qui nous ont permis d'améliorer ce document.

## 6.7. Bibliographie

[ANT 01] ANTONIADOU A., BESSE P., FOUGÈRES A.-L., LE GALL C., STEPHENSON D., « L'Oscillation Atlantique Nord (NAO) et son influence sur le climat européen », *Revue de Statistique Appliquée*, vol. 49, p. 39-60, 2001.

[BES 86] BESSE P., RAMSAY J., « Principal component analysis of sampled curves », *Psychometrika*, vol. 51, p. 285-311, 1986.

[BES 92] BESSE P., « PCA stability and choice of dimensionality », *Statistics & Probability Letters*, vol. 13, p. 405-410, 1992.

[BES 96] BESSE P., CARDOT H., « Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1 », *Canadian Journal of Statistics*, vol. 24, p. 467-487, 1996.

[BES 97] BESSE P., CARDOT H., FERRATY F., « Simultaneous nonparametric regressions of unbalanced longitudinal data », *Computational Statistics and Data Analysis*, vol. 24, p. 255-270, 1997.

[BES 00] BESSE P., CARDOT H., STEPHENSON D., « Autoregressive forecasting of some functional climatic variations », *Scandinavian Journal of Statistics*, vol. 27, p. 673-688, 2000.

[BOS 91] BOSQ D., « Modelization, non-parametric estimation and prediction continuous time processes », ROUSSAS G., Ed., *Nonparametric Functional Estimation and Related Topics*, Nato, Asi series, p. 509-529, 1991.

[BOS 00] BOSQ D., *Linear processes in function spaces*, Springer, 2000.

[CAR 98] CARDOT H., « Convergence du lissage spline de la prévision des processus autorégressifs fonctionnels », *C. R. Acad. Sci. Paris, Série I, t. 326*, p. 755-758, 1998.

[CAR 99a] CARDOT H., FERRATY F., SARDA P., « Functional Linear Model », *Statist. & Prob. Letters*, vol. 45, p. 11-22, 1999.

[CAR 99b] CARDOT H., FERRATY F., SARDA P., Spline Estimators for the Functional Linear Model : Consistency, Applications and Splus implementation., Rapport n°1, INRA Toulouse, Biometry and Artificial Intelligence, 1999.

[CAR 00a] CARDOT H., « Nonparametric estimation of the smoothed principal components analysis of sampled noisy functions », *Journal of Nonparametric Statistics*, vol. 12, p. 503-538, 2000.

[CAR 00b] CARDOT H., FAIVRE R., GOULARD M., VIALARD H., Estimations de l'occupation des sols à partir de l'évolution temporelle de images du capteur Végétation de Spot, Rapport, INRA Toulouse, Biométrie et Intelligence Artificielle, 2000.

[CAR 02a] CARDOT H., FERRATY F., MAS A., SARDA P., « Testing Hypotheses in the Functional Linear Model », *Scandinavian Journal of Statistics, to appear*, 2002.

[CAR 02b] CARDOT H., SARDA P., Estimation in Generalized Linear Models for Functional Data via Penalized Likelihood, Rapport n° 1, Publications du Laboratoire de Statistique et Probabilités, Toulouse, 2002.

[CAU 86] CAUSSINUS H., « Models and uses of principal component analysis », DE LEEUW J. et al., Eds., *Multidimensional Data Analysis*, DSWO Press, p. 149-170, 1986.

[DAM 02] DAMON J., GUILLAS S., « The inclusion of exogenous variables in functional autoregressive ozone forecasting », *Environmetrics, à paraître*, vol. 0, p. 000-000, 2002.

[DAU 76] DAUXOIS J., POUSSE A., Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique, PhD thesis, Thèse d'État, Université Toulouse III, 1976.

[DEB 78] DE BOOR C., *A Practical Guide to Splines*, Springer Verlag, 1978.

[DEV 74] DEVILLE J., *Méthodes statistiques et numériques de l'analyse harmonique*, Ann Insee, 15., 1974.

[DIE 93] DIERCKX P., *Curve and Surface Fitting with Splines*, Clarendon Press. Oxford, 1993.

[DO 99] DO K., KIRK K., « Discriminant Analysis of Event-Related Potential Curves Using Smoothed Principal Components », *Biometrics*, vol. 55, p. 174-181, 1999.

[ECO 91] ECOSTAT, *Banque de données statistiques pour l'enseignement*, C.R.D.P. de Montpellier, Unité de la documentation Statistique, 1991.

[FAH 94] FAHRMEIR L., TUTZ G., *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Verlag, 1994.

[FER 02a] FERRATY F., GOIA A., VIEU P., « Functional Nonparametric Model for Time Series : A Fractal Approach for Dimension Reduction », *TEST, to appear*, vol. 00, p. 000-000, 2002.

[FER 02b] FERRATY F., VIEU P., « The Functional Nonparametric Model and Application to Spectrometric Data », *Computational Statistics, to appear*, vol. 00, p. 000-000, 2002.

[FRA 93] FRANK I., FRIEDMAN J., « A Statistical View of Some Chemometrics Regression Tools », *Technometrics*, vol. 35, p. 109-148, 1993.

[GOU 98] GOUTIS C., « Second-derivative functional regression with applications to near infra-red spectroscopy », *Journal of the Royal Statistical Society, B*, vol. 60, p. 103-114, 1998.

[GRE 94] GREEN P., SILVERMAN B., *Nonparametric regression ans generalized linear model : a roughness penalty approach*, Chapman and Hall, 1994.

[HAS 93] HASTIE T., MALLOWS C., « A discussion of "A Statistical View of Some Chemometrics Regression Tools" by I.E. Frank and J.H. Friedman », *Technometrics*, vol. 35,

p. 140-143, 1993.

[HAS 95] HASTIE T., BUJA A., TIBSHIRANI R., « Penalized discriminant analysis », *Annals of Statistics*, vol. 23, p. 73-102, 1995.

[KNE 00] KNEIP A., LI X., MACGIBBON B., RAMSAY J. O., « Curve registration by local regression », *Canadian Journal of Statistics*, vol. 28, p. 19-30, 2000.

[LEU 93] LEURGANS S., MOYEED R., SILVERMAN B., « Canonical Correlation Analysis when the Data are Curves », *Journal of the Royal Statistical Society, B*, vol. 55, p. 725-740, 1993.

[MAR 96] MARX B., EILERS P., « Generalized Linear Regression on Sampled Signals with Penalized Likelihood », FORCINA A., MARCHETTI G., R. H., GALMACCI G., Eds., *11th International workshop on Statistical Modelling*, 1996.

[MAR 99] MARX B., EILERS P., « Generalized Linear Regression on Sampled Signals and Curves : A *P*-Spline Approach », *Technometrics*, vol. 41, p. 1-13, 1999.

[MAS 00] MAS A., Estimation d'opérateurs de corrélation de processus fonctionnels : lois limites, tests, déviations modérées, PhD thesis, Université Paris 6, 2000.

[MOU 95] MOURID T., *Contribution à la statistique des processus autorégressifs à temps continu*, Thèse d'Etat, Université Paris VI, 1995.

[PRE 02] PREDA C., SAPORTA G., « Régression PLS sur un processus stochastique », *Revue de Statistique Appliquée*, vol. 50, p. 27-45, 2002.

[PUM 92] PUMO B., Estimation et prévision de processus fonctionnels autorégressifs : Application aux processus à temps continu, PhD thesis, Université Paris 6, 1992.

[RAM 91] RAMSAY J., DALZELL C., « Some tools for functional data analysis », *Journal of the Royal Statistical Society, B*, vol. 53, p. 539-572, 1991, with discussion.

[RAM 96] RAMSAY J., « Principal differential analysis : data reduction by differential operators », *Journal of the Royale Statistical Society Ser. B*, vol. 58, p. 495–508, 1996.

[RAM 97] RAMSAY J., SILVERMAN B., *Functional Data Analysis*, Springer-Verlag, 1997.

[RAM 98] RAMSAY J. O., LI X., « Curve registration », *Journal of the Royal Statistical Society, Series B,*, vol. 60, p. 351–363, 1998.

[RAM 00] RAMSAY J., « Differential equation models for statistical functions », *Canadian Journal of Statistics*, vol. 28, n° 1, p. 225–240, 2000.

[RIC 01] RICE J., WU C., « Nonparametric mixed effects models for unequally sampled noisy curves », *Biometrics*, vol. 57, p. 253-259, 2001.

[SIL 96] SILVERMAN B., « Smoothed functional principal components analysis by choice of norm », *The Annals of Statistics*, vol. 24, p. 1-24, 1996.

[SPL 97] S-PLUS, *S-PLUS 4 Guide to statistics*, Mathsoft, 1997.

[WAH 90] WAHBA G., *Spline Models for Observational Data*, SIAM, 1990.

[YAO 00] YAO A. F., FERRÉ L., Functional Sliced Inverse Regression analysis, Rapport n° 14, Publications du Laboratoire de Statistique et Probabilités, Toulouse, 2000.