Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1

Philippe C. BESSE et Hervé CARDOT
Laboratoire de Statistique et Probabilités, UMR CNRS
Université Paul Sabatier, 31062 Toulouse cedex, FRANCE
Email : besse@cict.fr

Résumé

Nous présentons dans cet article une méthode d'approximation et de prévision, sur un intervalle de temps, de processus à temps continu. Pour cela, nous utilisons les processus ARH(1), définis par Bosq (1991). Nous abordons le problème, important dans la pratique, de la dicrétisation des trajectoires observées en les approximant par des fonctions spline. Pour des paramètres de lissage bien choisis au sens d'un critère de type validation croisée, nous obtenons de bons résultats, aussi bien sur des simulations que sur des données réelles. Ce modèle est enfin appliqué à la prévision du trafic routier et comparé avec un modèle SARIMA.

Abstract

We present a method to approximate and forecast, on an entire interval, a continuous time process. For this purpose, we use the modelization of ARH(1) processes, defined by Bosq (1991). We deal with the practical problem of the discretization of the observed trajectories and approximate them by means of splines functions. We show by simulations that for "well chosen" smoothing parameters, good prediction can be obtained compared to the "predictable" part of the process. Finally, we apply this model to forecast road traffic and compare it to a SARIMA model.

Key words : functional prediction, ARH(1) process, smoothing spline, principal component analysis.
AMS 1991 subject classification : 62H25, 62G07, 62M20.

1 Introduction

Cet article s'intéresse à la prévision d'un processus présentant naturellement une composante périodique. C'est le cas des données étudiées dans la section 5, issues de l'observation du trafic horaire autoroutier à un poste de péage. On se propose de prévoir le trafic d'un jour donné à partir de celui des jours identiques des semaines précédentes. Mais, plutôt que de prévoir séquentiellement la série des trafics horaires comme le ferait une modélisation SARIMA, on s'intéresse à une *prévision globale* sur la période considérée.

Notre objectif est donc la prévision d'un processus $(X_t)_{t \in \mathbf{R}}$ à temps continu, pas nécessairement stationnaire, sur un intervalle de temps δ, dont on fixe la valeur $\delta = 1$ sans perte de généralité. On définit alors, à partir de (X_t), un processus à temps discret $(Y_n)_{n \in \mathbb{Z}}$, 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 $\delta = 1$.

Le cadre théorique de cette approche a été proposé par Bosq (1991), qui s'intéresse à des processus fonctionnels (Y_n) autorégressifs 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). Dans ce cas, l'estimation de l'espérance conditionnelle $E(Y_{i+1}|Y_i, Y_{i-1}, \ldots)$ nécessaire à la prévision se ramène à celle de l'opérateur ρ :

$$E(Y_{i+1}|Y_i, Y_{i-1}, \ldots) - a = \rho(Y_i - a) \;, \quad i \in \mathbb{Z}, \tag{1}$$

appelé opérateur d'autocorrélation et où la fonction a représente la moyenne du processus $(Y_i)_{i \in \mathbb{Z}}$.

Notons Γ (resp. Δ^*) l'opérateur de covariance (resp. de covariance croisée) du processus, qui sont définis par :

$$\begin{aligned} \forall\, f \in L^2[0,1], \qquad \Gamma f(t) &= \int_{[0,1]} R(s,t) f(s) ds \;,\; t \in [0,1], \\ \Delta^* f(t) &= \int_{[0,1]} R(s+1,t) f(s) ds \;,\; t \in [0,1] \end{aligned}$$

où $R(s,t)$ est la covariance entre X_s et X_t. Ces opérateurs vérifient alors la relation suivante :

$$\Delta^*(Y_i - a) = \Gamma(E(Y_{i+1}|Y_i, Y_{i-1}, \ldots) - a) \tag{2}$$

et, sous l'hypothèse que (Y_i) est du second ordre, possèdent un spectre discret (ils sont compacts et même de Hilbert-Schmidt).

L'opérateur Γ n'étant pas inversible, Bosq (1991) propose la démarche suivante afin de construire un estimateur de ρ :

1. analyse spectrale de l'opérateur Γ fournie par les solutions de l'équation intégrale
$$\int_{[0,1]} R(s,t) f(s) ds = \lambda f(t) \;,\; t \in [0,1] \;\; ;$$

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 Γ associées aux q plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_q > 0$;

3. inversion de la restriction de Γ à H_q ;

4. approximation de rang fini d'un estimateur de ρ à partir de (2).

La mise en œuvre pratique d'une telle démarche pose ensuite deux problèmes : le processus est observé à p instants de discrétisation $\{t_1, \ldots, t_p\}$ sur un nombre fini n d'intervalles de temps. On construit alors des estimateurs à partir de l'analyse spectrale de la matrice de covariance empirique. Le problème de l'échantillonnage est traité par Bosq (1991), qui étudie les propriétés asymptotiques des estimateurs et montre leur convergence.

La principale contribution de cet article est la prise en compte du problème de la discrétisation des trajectoires *conjointement* à la réduction de dimension. Nous proposons et comparons ainsi différents types d'approximation des trajectoires reposant sur des hypothèse classiques :

- la partie imprévisible du processus, l'innovation, est associée à la partie irrégulière (le bruit) des trajectoires;
- la partie prévisible des trajectoires est supposée régulière comme il est habituel en estimation fonctionnelle de la régression par lissage spline.

La démarche ainsi proposée, qui consiste à étudier n courbes observées à p instants de discrétisation, présente alors une forte analogie avec d'autres travaux portant sur l'analyse en composantes principales (ACP) de courbes, Besse & Ramsay (1986), Bouhaddou et coll. (1987), Ramsay & Dalzell (1991), Rice & Silverman (1991), Besse & Pousse (1992), Besse et coll. (1995), qui poursuivent également un objectif de réduction de dimension.

La section 2 propose une solution au problème de prévision en appliquant les résultats de Besse et coll. (1995), qui consistent à intégrer des contraintes de régularité ou de lissage dans un problème de réduction de rang. Dans ce cas, l'utilisation de splines hybrides (Kelly & Rice, 1990), c'est-à-dire l'approximation des splines de lissage par des B-splines, permettrait même de considérer un processus dont les instants d'observation sont différents d'une période à l'autre. La section 3 décrit d'autres prédicteurs pouvant être considérés en fonction de la procédure d'approximation des trajectoires utilisée :

- approximation linéaire, Pumo (1992);
- lissage puis réduction de rang, Besse & Pousse (1992);
- recherche de fonctions propres lisses, Rice & Silverman (1991).

La section 4 compare différentes prévisions entre elles sur des données simulées. Enfin, à la section 5, ces méthodes sont appliquées à la prévision du trafic autoroutier et comparées avec un modèle SARIMA habituellement utilisé pour ce type de données.

2 Approximation spline avec contrainte de rang

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 (1991) et Pumo (1992) repose sur deux points importants :

- Une estimation de l'opérateur ρ de rang fini;

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

Nous proposons de remplacer l'interpolation linéaire par une estimation nonparamétrique des trajectoires avant d'entamer l'approximation de l'opérateur de covariance Γ. 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 nonparamétrique 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 résolu par Besse et coll. (1995) dans le cadre de l'estimation fonctionnelle simultanée de plusieurs observations d'un processus, est donc intégrée dans le modèle de prévision décrit ci-après.

2.1 Lissage spline

L'estimation nonparamétrique par lissage spline a donné lieu à une importante littérature dont Wahba (1990) et Green & Silverman (1994) donnent un aperçu. Plaçons-nous dans le cadre du modèle suivant :

$$y_j = f(t_j) + \epsilon_j \ ; \quad E(\epsilon_j) = 0,\ E(\epsilon_i \epsilon_j) = \sigma^2 \delta_{ij},\ i, j = 1, ..., p$$
$$0 \leq t_1 < t_2 < ... < t_p \leq 1.$$

où f est supposée régulière et lisse.

L'hypothèse de régularité est satisfaite en imposant à la fonction f d'appartenir à un espace de Sobolev

$$\begin{aligned} W^m[0,1] \quad = \quad & \Big\{ f \ : \ [0,1] \to \mathbb{R} \ ; \ \ f,\ Df,\ D^{m-1}f \text{ absolument continues,} \\ & D^m f \in L^2[0,1] \Big\}. \end{aligned}$$

La contrainte imposant à la courbe d'être lisse est définie par :

$$\int_0^1 (D^m f(t))^2 dt = \|D^m f\|^2_{L^2} < \mathrm{c}, \quad \mathrm{c} \ \in \mathbb{R}_+.$$

L'opérateur différentiel D^m, m appartenant à $\mathbb{N}$, permet de mesurer “combien la courbe est lisse” en définissant une semi-norme sur W^m. Pour la suite, nous avons fixé $m = 2$; par conséquent, les trajectoires sont lissées par des splines cubiques.

L'estimation spline $\hat{f}$ de la fonction f est alors solution du problème d'optimisation suivant :

$$\min \left\{ \frac{1}{p} \sum_{j=1}^{p} (f(t_j) - y(t_j))^2; \ \left\| D^2 f \right\|^2_{L^2} < \mathrm{c} \ (\mathrm{c} \ \in \mathbb{R}_+),\ f \in W^2 \right\}. \tag{3}$$

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

$$\min \left\{ \frac{1}{p} \sum_{j=1}^{p} (f(t_j) - y(t_j))^2 + \ell \left\| D^2 f \right\|_{L^2}^2 ; f \in W^2 \right\}. \tag{4}$$

Le paramètre de lissage ℓ permet d'effectuer un arbitrage entre la fidélité aux données ($\ell = 0$) et la régularité de la solution ($\ell \to +\infty$). La solution $\widehat{f}$ de ce problème est une fonction polynômiale par morceaux qui interpole aux points de discrétisation les valeurs contenues dans le vecteur [1]

$$\widehat{\mathbf{y}} = \mathbf{A}(\ell)\mathbf{y},$$

où $\mathbf{A}(\ell)$ est la matrice de "lissage" (Hat-matrix) associée au paramètre ℓ. On peut montrer que $\widehat{f}$ est un élément d'un sous-espace de W^2 de dimension finie p. Soit $\{h_1, \ldots, h_p\}$ une base de cet espace et notons $\mathbf{K}_1$ la matrice $(p \times p)$ permettant d'obtenir, à partir du vecteur des observations lissées $\mathbf{A}(\ell)\mathbf{y}$, le vecteur des coordonnées $\mathbf{K}_1\mathbf{A}(\ell)\mathbf{y}$ de la fonction $\widehat{f}$ dans cette base. De plus, notons $\mathbf{K}$ la matrice $(p \times p)$ vérifiant $\left\| D^2 \hat{f} \right\|_{L^2} = \|\widehat{\mathbf{y}}\|_K$. C'est, pour la base considérée, la matrice associée à la semi-norme induite de W^2 sur le sous-espace spline. On pourra se reporter à Besse & Ramsay (1986) et à Ramsay & Dalzell (1991) pour le calcul pratique des matrices $\mathbf{K}_1$ et $\mathbf{K}$.
Définissons enfin $\mathbf{C}$ la matrice $(p \times p)$ du produit scalaire de la base des fonctions splines dans $L^2[0,1]$:

$$[\mathbf{C}]_{ij} = \int_0^1 h_i(t) h_j(t) dt \; .$$

Il est alors facile de vérifier que :

$$< \hat{f}_1, \hat{f}_2 >= \int_0^1 \hat{f}_1(t) \hat{f}_2(t) \; dt = \mathbf{y}_1' \mathbf{A}(\ell) \mathbf{K}_1 \mathbf{C} \mathbf{K}_1 \mathbf{A}(\ell) \mathbf{y}_2, \tag{5}$$

où $\hat{f}_1$ (resp. $\hat{f}_2$) est l'estimateur spline calculé à partir du vecteur d'observations $\mathbf{y}_1$ (resp. $\mathbf{y}_2$). Cette relation entre le produit scalaire dans $L^2[0,1]$ et celui entre les observations nous permettra par la suite d'exprimer matriciellement les opérateurs de covariance et ainsi d'approcher leur analyse spectrale par une simple diagonalisation de matrice.

2.2 Modèle

On observe n réalisations d'un processus ARH(1), $(Z_i)_{i \in 1,..,n}$ en p points, $\{t_1, t_2, .., t_p\}$ de l'intervalle $[0,1]$. Notons $\mathbf{y}_i$ pour $i \in \{1, .., n\}$ les n vecteurs de $\mathbb{R}^p$ qui sont des observations discrètes du processus (Z_i).

Nous effectuons une approximation des trajectoires Z_i en supposant que la partie prévisible des trajectoires du processus est lisse et appartient à un sous-espace H_q de fonctions régulières de dimension finie q. Cela revient à supposer que l'espace image de $L^2[0,1]$ par l'opérateur ρ est un espace de fonctions lisses et régulières. Notre objectif, qui est la prévision de Z_{n+1}, nécessite d'estimer ρ, Z_n et a, la moyenne du processus.

1. Les **matrices** et les **vecteurs** sont notés en gras pour les distinguer des fonctions.

2.3 Estimation avec lissage sous contrainte de rang

Les données sont rangées dans une matrice $\mathbf{Y}$ de taille $(n \times p)$. À chaque trajectoire est associé un vecteur ligne $\mathbf{y}_i$ de la matrice $\mathbf{Y}$. Notons $\mathbf{X}$ la matrice des trajectoires centrées; la ligne numéro i est définie par $\mathbf{x}_i = \mathbf{y}_i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_j$.

On se propose d'estimer les trajectoires du processus en résolvant le problème d'optimisation suivant :

$$\min_{f_i}\left\{\frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}(\mathbf{y}_i(t_j) - f_i(t_j))^2\ ;\forall i,\ f_i \in H_q,\ \ \dim(H_q) = q,\ \left\|D^2 f_i\right\|_{L^2}^2 < c\right\}, \quad (6)$$

qui se ramène à :

$$\min_{f_i}\left\{\frac{1}{np}\sum_{i=1}^{n}\left(\sum_{j=1}^{p}(\mathbf{y}_i(t_j) - f_i(t_j))^2 + \ell\left\|D^2 f_i\right\|_{L^2}^2\right);\forall i,\ f_i \in H_q,\ \ \dim(H_q) = q\right\}. \quad (7)$$

Cette méthode permet d'anticiper la réduction nécessaire du rang de l'opérateur ρ en imposant à la fois aux fonctions f_i d'être lisses et d'appartenir à un sous-espace de dimension finie q.
Notons $\mathbf{D} = \frac{1}{n}\mathbf{I}_n$ où $\mathbf{I}_n$ est la matrice identité de dimension $n \times n$ et définissons la matrice symétrique

$$\mathbf{S}_1 = \mathbf{A}(\ell)^{1/2}\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{A}(\ell)^{1/2}.$$

Les éléments propres $(\lambda_1, \mathbf{v}_1), \ldots, (\lambda_p, \mathbf{v}_p)$ de $\mathbf{S}_1$ sont rangés selon l'ordre décroissant des valeurs propres.

Proposition 2.1 *La solution du problème (7) est fournie par les interpolations splines $\widehat{f}_i$ aux points contenus dans les vecteurs $\widehat{\mathbf{y}}_i$:*

$$\widehat{\mathbf{y}_i} \;=\; \mathbf{P}_q\left(\mathbf{X}\mathbf{A}(\ell)\mathbf{x}_i\right) + \widehat{\mathbf{a}}, \quad (8)$$

où

$$\widehat{\mathbf{a}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{A}(\ell)\mathbf{y}_i = \mathbf{A}(\ell)\bar{\mathbf{y}},$$

et où $\mathbf{P}_q$ est le projecteur $\mathbf{A}(\ell)^{-1}$-orthogonal sur le sous-espace engendré par les vecteurs $\mathbf{A}(\ell)^{1/2}\mathbf{v}_1, \ldots, \mathbf{A}(\ell)^{1/2}\mathbf{v}_q$.

Démonstration
Notons $\mathbf{z}_i$ le vecteur de $\mathbb{R}^p$ associé à f_i; soient aussi $\mathbf{z}_i(j) = f_i(t_j)$ et $\bar{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i$. On définit enfin la matrice centrée $\mathbf{Z}$ $(n \times p)$ dont les lignes sont $(\mathbf{z}_i - \bar{\mathbf{z}})'$.
Le critère à minimiser se décompose de la façon suivante :

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(\|\mathbf{y}_i - \mathbf{z}_i\|^2_{\mathbf{I}_p} + \ell\,\|\mathbf{z}_i\|^2_{\mathbf{K}}\right) \;&=\; \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - (\mathbf{z}_i - \bar{\mathbf{z}})\|^2_{\mathbf{I}_p} + \ell\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{z}_i - \bar{\mathbf{z}}\|^2_{\mathbf{K}} + \\ &+\; \|\bar{\mathbf{y}} - \bar{\mathbf{z}}\|^2_{\mathbf{I}_p} + \ell\,\|\bar{\mathbf{z}}\|^2_{\mathbf{K}}.\end{aligned}$$

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

$$\widehat{\mathbf{a}} = \mathbf{A}(\ell)\bar{\mathbf{y}},$$

tandis que les deux premiers termes nous amènent à résoudre :

$$\min_{\mathbf{Z}} \left\{ \|\mathbf{Z} - \mathbf{X}\|^2_{\mathbf{I}_p,\mathbf{D}} + \ell \, \|\mathbf{Z}\|^2_{\mathbf{K},\mathbf{D}} \, ; \; \mathrm{rang}(\mathbf{Z}) = q (q < p) \right\} \tag{9}$$

où $\|\mathbf{Z}\|^2_{\mathbf{K},\mathbf{D}} = \mathrm{tr}\mathbf{Z}'\mathbf{D}\mathbf{Z}\mathbf{K}$.

Notons $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{A}(\ell)$ la matrice des lignes “lissées” de $\mathbf{X}$, de sorte que

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

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 $\widetilde{\mathbf{X}}$. Elle est obtenue par la décomposition en valeurs singulières (DVS) de $\mathbf{X}\mathbf{A}(\ell)$ relative aux métriques $\mathbf{A}(\ell)^{-1}$ et $\mathbf{D}$:

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

$$\text{où} \left\{ \begin{array}{lcl} \mathbf{X}\mathbf{A}(\ell)\mathbf{X}'\mathbf{D}\widetilde{\mathbf{U}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{L}} & \text{et} & \widetilde{\mathbf{U}}'\mathbf{D}\widetilde{\mathbf{U}} = \mathbf{I}_p, \\ \mathbf{A}(\ell)\mathbf{X}'\mathbf{D}\mathbf{X}\widetilde{\mathbf{V}} = \widetilde{\mathbf{V}}\widetilde{\mathbf{L}} & \text{et} & \widetilde{\mathbf{V}}'\mathbf{A}(\ell)^{-1}\widetilde{\mathbf{V}} = \mathbf{I}_p. \end{array} \right.$$

Cette décomposition en valeurs singulières généralisée peut être obtenue plus facilement par celle de $\mathbf{X}\mathbf{A}(\ell)^{1/2}$ relative à $\mathbf{I}_p$ et $\mathbf{D}$:

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

$$\text{où} \left\{ \begin{array}{lcl} \mathbf{X}\mathbf{A}(\ell)^{1/2}\mathbf{A}(\ell)^{1/2}\mathbf{X}'\mathbf{D}\mathbf{U} = \mathbf{U}\mathbf{L} & \text{et} & \mathbf{U}'\mathbf{D}\mathbf{U} = \mathbf{I}_p, \\ \mathbf{A}(\ell)^{1/2}\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{A}(\ell)^{1/2}\mathbf{V} = \mathbf{V}\mathbf{L} & \text{et} & \mathbf{V}'\mathbf{V} = \mathbf{I}_p. \end{array} \right.$$

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 DVS de $(\mathbf{X}\mathbf{A}(\ell)^{1/2}, \mathbf{I}_p, \mathbf{D})$ conduit à l'analyse spectrale de la matrice $\mathbf{S}_1 = \mathbf{A}(\ell)^{1/2}\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{A}(\ell)^{1/2}$. Les trajectoires discrètes du processus sont projetées sur le sous-espace engendré par les vecteurs

$$\tilde{\mathbf{v}}_i = \mathbf{A}(\ell)^{1/2}\mathbf{v}_i, \quad i = 1, 2, ..., q.$$

$\square$

Remarque : Cette méthode d'approximation des trajectoires peut être adaptée au cas où la discrétisation varie d'une trajectoire à l'autre en utilisant des splines hybrides, c'est-à-dire une approximation des splines de lissage dans une base de B-splines (voir Besse et coll. 1995).

2.4 Prévision

On définit alors, en plongeant $W^2[0,1]$ dans l'espace $L^2[0,1]$, des estimateurs des opérateurs de covariance par :

$$\widehat{\Gamma} f(t) = \int_0^1 \widehat{R}(s,t) f(s) ds, \quad f \in L^2[0,1]\,, \tag{10}$$

$$\widehat{\Delta^*} f(t) = \int_0^1 \widehat{R}(s+1,t) f(s) ds, \quad f \in L^2[0,1]\,, \tag{11}$$

où $\widehat{R}(s,t)$ est l'estimateur empirique de la covariance obtenu à partir des trajectoires $\hat{\mathbf{y}}_i$. À partir de la relation (5), on déduit une représentation matricielle de ces opérateurs :

$$\widehat{\boldsymbol{\Gamma}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}_q \mathbf{A}(\ell) \mathbf{x}_i \mathbf{x}_i' \mathbf{A}(\ell) \mathbf{P}'_q \mathbf{M}, \tag{12}$$

$$\widehat{\boldsymbol{\Delta}^*} = \frac{1}{n-1} \sum_{i=2}^{n-1} \mathbf{P}_q \mathbf{A}(\ell) \mathbf{x}_{i-1} \mathbf{x}_i' \mathbf{A}(\ell) \mathbf{P}'_q \mathbf{M}, \tag{13}$$

où $\mathbf{M} = \mathbf{K}_1 \mathbf{C} \mathbf{K}_1$.

Enfin, en utilisant (2), on définit un estimateur de l'opérateur d'autocorrélation en posant :

$$\widehat{\rho} = \widehat{\Gamma}^{-1} \widehat{\Delta^*}, \tag{14}$$

où $\widehat{\Gamma}^{-1}$ est l'inverse de $\widehat{\Gamma}$ dans l'espace H_q. Il est alors possible d'effectuer une prévision de la trajectoire de Z_{n+1} :

$$\tilde{Z}_{n+1} = \widehat{\rho}(\hat{y}_n - \hat{a}) + \hat{a}. \tag{15}$$

3 Autres méthodes d'approximation

Nous présentons dans cette partie d'autres techniques qui permettent d'estimer les opérateurs de covariance du processus observé.

3.1 Interpolation linéaire

Pumo (1992) effectue une interpolation linéaire des trajectoires entre les points de discrétisation puis estime, en utilisant (10) et (11), les opérateurs de covariance. Un estimateur de la moyenne est ainsi obtenu en interpolant la moyenne empirique. D'autre part, si l'on note $\{e_1, \ldots, e_p\}$ le système de fonctions linéaires utilisé pour l'interpolation et $\mathbf{E}$ la matrice de dimension $p \times p$ vérifiant :

$$\int_0^1 \widehat{f}_1(t) \widehat{f}_2(t) dt = \mathbf{y}_1' \mathbf{E} \mathbf{y}_2,$$

alors l'opérateur de covariance admet pour représentation matricielle :

$$\widehat{\mathbf{\Gamma}} = \mathbf{X}'\mathbf{DXE}.$$

Cette solution ne comporte pas de paramètre de lissage et est, par conséquent, moins flexible. Nous verrons par la suite qu'elle fournit de moins bonnes prévisions.

3.2 Lissage spline

On effectue une régression non paramétrique avec interpolation ou lissage spline de chaque trajectoire observée $\mathbf{y}_i$, *puis* on estime $\widehat{\Gamma}$ et $\widehat{\Delta^*}$. Cette méthode a été présentée par Besse & Pousse (1992) dans le cas de "courbes" indépendantes entre-elles pour l'estimation de l'opérateur de covariance en vue de l'approximation spline de l'ACP. Les trajectoires lisses du processus sont estimées à partir de l'équation (4).

Proposition 3.1 *L'estimation des trajectoires s'effectue à partir de l'analyse spectrale de la matrice S_2 définie par :*

$$\mathbf{S}_2 = \mathbf{A}(\ell)\mathbf{X}'\mathbf{DXA}(\ell)\mathbf{M},$$

et la moyenne discrétisée est estimée par $\hat{\mathbf{a}}$.

Démonstration
Le vecteur de la $i$ème trajectoire lisse et discrète est défini par $\widetilde{\mathbf{y}}_i = \mathbf{A}(\ell)\mathbf{y}_i$. La moyenne lisse et discrète du processus est estimée par

$$\overline{\widetilde{\mathbf{y}}} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{y}}_i = \mathbf{A}(\ell)\bar{\mathbf{y}},$$

tandis que l'opérateur de covariance admet la représentation matricielle :

$$\begin{aligned} \widehat{\mathbf{\Gamma}} &= \frac{1}{n}\sum_{j=1}^{n}\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j'\mathbf{M} \\ &= \mathbf{A}(\ell)\mathbf{X}'\mathbf{DXA}(\ell)\mathbf{M}, \end{aligned}$$

où $\widetilde{\mathbf{x}}_j$ est la trajectoire lissée et centrée $\mathbf{A}(\ell)\mathbf{x}_i$. □

On en déduit les opérateurs de covariance empiriques à partir des équations (10) et (11). On peut alors effectuer une prévision de Z_{n+1} en utilisant (14).

3.3 Estimation de fonctions propres lisses

On estime directement, sous contrainte de régularité, les fonctions propres de l'opérateur de covariance du processus. Cette méthode a été proposée par Rice & Silverman (1991) dans le cas indépendant. Elle consiste à chercher la solution du problème d'optimisation suivant :

$$\underset{f\in L^2[0,1]}{\mathrm{Max}} \left\{ \int_0^1 \Gamma f(t) f(t) dt \; - \ell \left\| D^2 f \right\|_{L^2}^2 ; \|f\|_{L^2} = 1 \right\} \tag{16}$$

en itérant sous contrainte d'orthonormalité. En d'autres termes, on recherche f_j solution du problème avec la contrainte additionnelle

$$\int_0^1 f_j(t) f_i(t) dt = \delta_{ij}, \; i \leq j.$$

Proposition 3.2 *On résout le problème (16) en effectuant l'analyse spectrale de la matrice :*

$$\mathbf{S}_3 = \mathbf{X}^{'}\mathbf{DXM} - \ell\mathbf{K}.$$

Les solutions sont les vecteurs propres de $\mathbf{S}_3$ *associés aux plus grandes valeurs propres, la matrice* $\mathbf{K}$ *étant définie en (2.1).*

Démonstration
La moyenne discrète du processus est estimée de la même façon que pour la méthode d'approximation spline. On mesure la régularité des fonctions discrètes $\mathbf{v}_i$ par $\mathbf{v}_i^{'}\mathbf{K}\mathbf{v}_i$. Le problème d'optimisation posé par Rice & Silverman (1991) se traduit alors en terme matriciel par :

$$\text{Max}\ \{\mathbf{v}^{'}\mathbf{X}^{'}\mathbf{DXMv} - \ell\mathbf{v}^{'}\mathbf{K}\mathbf{v};\ \mathbf{v} \in \mathbb{R}^p\}$$

en itérant sous contrainte d'orthonormalité. La solution de ce problème est fournie par l'analyse spectrale de la matrice $\mathbf{X}^{'}\mathbf{DXM} - \ell\mathbf{K}$. □

Un estimateur des trajectoires du processus est alors

$$\hat{\mathbf{y}}_i = \mathbf{P}_q(\mathbf{x}_i) + \hat{\mathbf{a}}\ ,$$

$\mathbf{P}_q$ étant le projecteur orthogonal sur les q premiers vecteurs propres de $\mathbf{S}_3$. On estime ensuite $\widehat{\Delta^*}$ en utilisant l'équation (11); on peut alors effectuer une prévision de Z_{n+1} à partir de (14).

Des propriétés de convergence asymptotique des estimateurs peuvent être obtenues sans trop de difficultés en combinant les résultats de Besse (1991) et Bosq (1991) pour l'approximation spline et en combinant ceux de Pezzulli & Silverman (1993) et Bosq (1991) pour l'estimation des fonctions propres lisses. En ce qui concerne l'approximation par interpolation linéaire, Pumo (1992) a montré la convergence en probabilité de $\widehat{\rho}$ vers ρ, si les trajectoires sont suffisamment régulières, lorsque le pas de discrétisation tend vers 0 et le nombre de trajectoires observées tend vers l'infini.

4 Comparaison sur des simulations

Nous simulons quelques processus $ARH(1)$ et comparons, pour les différentes méthodes étudiées précédemment, la qualité des prévisions fournies. Deux techniques de simulation ont été employées.

4.1 Par construction d'un opérateur d'autocorrélation ρ

La démarche suivie par Pumo (1992) pour simuler un processus $ARH(1)$ repose sur les étapes suivantes :

1. Simulation d'un bruit blanc fonctionnel, en l'occurrence n trajectoires d'un mouvement brownien indépendantes sur $[0, 1]$;

2. Définition de l'opérateur ρ par l'intermédiaire de ses fonctions propres et valeurs propres associées;

3. Simulation en utilisant l'équation de récurrence (1) caractérisant le processus.

4.1.1 Simulation des mouvements browniens

Pumo (1992) utilise la décomposition de Karhunen-Loeve pour simuler les trajectoires de mouvements browniens indépendants sur $[0, 1]$. Cette méthode est fastidieuse à programmer; c'est pourquoi nous lui avons préféré une méthode beaucoup plus rapide qui repose sur le fait que l'accroissement entre 2 dates t_1 et t_2, $t_2 > t_1$, d'un mouvement brownien B_t suit une loi normale centrée, indépendante de B_{t_1} et de variance $|t_1 - t_2|$. Pour simuler un mouvement brownien sur $[0, 1]$ en p points de discrétisation $t_1, t_2, ..., t_p$ il suffit donc de simuler p variables aléatoires W_i, $i = 1, 2, ..., p$, gaussiennes, centrées et de variance $\mathrm{Var}(W_i)=t_i - t_{i-1}$, avec $W_0 = 0$. Alors

$$B_{t_i} = \sum_{j=1}^{i} W_j.$$

En itérant n fois cette procédure, nous avons simulé n bruits blancs fonctionnels, notés $\epsilon_1, \epsilon_2, ..., \epsilon_n$.

A priori, on pourrait penser qu'il y a une contradiction entre le choix de l'espace $W^2[0, 1]$ et le fait que les trajectoires du bruit blanc appartiennent à $C[0, 1]$. Cependant les q premières fonctions propres de l'opérateur de covariance du processus appartiennent à $C^\infty[0, 1] \subset W^2[0, 1]$, ce qui reste en accord avec le modèle présenté en (2.2).

4.1.2 Définition de l'opérateur d'autocorrelation ρ

Notons Γ_ϵ l'opérateur de covariance d'un mouvement brownien. Cet opérateur vérifie :

$$\Gamma_\epsilon f(t) = \int_0^1 \min(s, t) f(s) ds \ , \ f \in L^2[0, 1], \ t \in [0, 1].$$

Les fonctions propres de Γ_ϵ sont

$$g_i(t) = \sqrt{2} \sin \left\{ \left(i - \frac{1}{2} \right) \pi t \right\}, \ t \in [0, 1]$$

associées aux valeurs propres

$$\lambda_i = \frac{1}{(i-\frac{1}{2})^2\pi^2}, \ i = 1, 2, \ldots$$

Définissons ensuite ρ par

$$\rho(g_i) = \beta_i g_i, \ \beta_i \in (-1, 1)$$

pour $i \leq \min(p, n)$. Nous imposons ainsi à ρ les mêmes vecteurs propres que Γ_ϵ et fixons les β_i pour définir ρ de manière explicite.

Il est alors possible de calculer l'opérateur de covariance Γ du processus qui est caractérisé par :

$$\Gamma g_i = \frac{\lambda_i}{1-\beta_i^2} g_i.$$

4.1.3 Simulation du processus

Pour utiliser l'équation (1), il suffit de considérer les coordonnées du bruit blanc défini sur $L^2[0,1]$ dans le système orthonormal $g_1, g_2, \ldots$ et d'appliquer l'équation de récurrence :

$$Z_i^j = \beta_j . Z_{i-1}^j + \epsilon_i^j,$$

où ϵ_i^k et Z_i^k sont les coordonnées de ϵ_i et Z_i sur g_k. Enfin un nouveau changement de coordonnées permet de revenir dans la base canonique.

Pour chaque simulation, nous avons calculé un rapport "Signal/Bruit" mesuré par :

$$\text{Rapport Signal/Bruit} = \frac{\text{trace } (\rho\Gamma\rho)}{\text{trace } (\Gamma)},$$

où

$$\text{trace } (\rho\Gamma\rho) = \sum_{j=1}^{q} \frac{\lambda_i \beta_i^2}{(1-\beta_i^2)}$$

et

$$\text{trace } (\Gamma) = \sum_{j=1}^{+\infty} \frac{\lambda_i}{(1-\beta_i^2)}.$$

Ce rapport s'interprète comme la part de variance d'une trajectoire qui peut s'expliquer par le passé, c'est-à-dire qui n'est pas due à l'innovation.

4.1.4 Choix des paramètres et critère de comparaison

La dimension du sous-espace de représentation et le paramètre de lissage optimisent le critère de validation-croisée suivant :

$$(R) \ = \underset{\ell,q}{\text{Min}} \left\{ \frac{1}{r_1} \sum_{j=0}^{r_1-1} \left\| \hat{\rho}_{(n,q,\ell)}(Y_{n+j}) - Y_{n+j+1} \right\|_{L^2}^2 \right\}. \qquad (17)$$

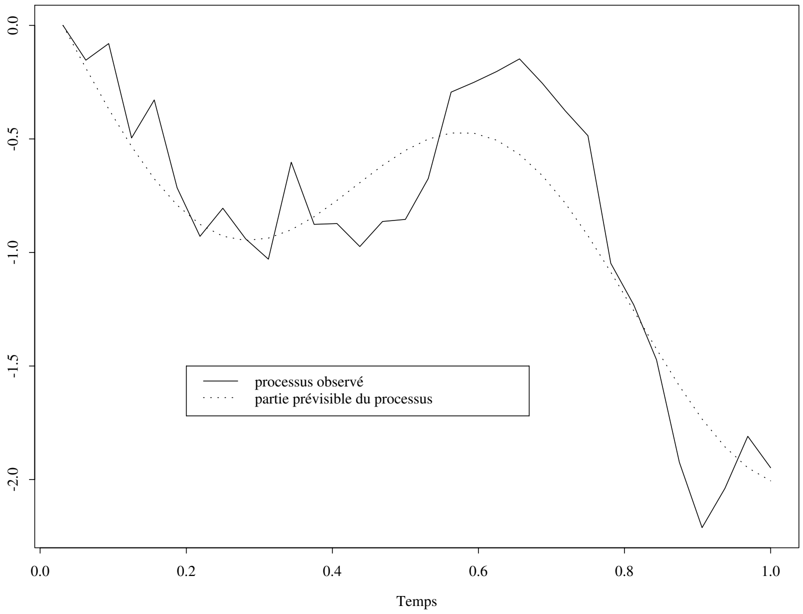

FIG. 1 - *Représentation d'une trajectoire simulée d'un processus ARH(1) et de sa partie prévisible (trajectoire privée de l'innovation); les données proviennent de la 2ème simulation.*

Nous avons utilisé n trajectoires pour estimer les opérateurs de covariance; puis, nous avons cherché les meilleures prévisions des r_1 trajectoires suivantes pour choisir ℓ et q. Les méthodes de prévision ont ensuite été comparées à l'aide du critère d'"erreur quadratique moyenne de prévision" de la partie prévisible du processus, que l'on note E.Q.M.P. et défini par :

$$\text{E.Q.M.P.} = \frac{1}{r_2} \sum_{j=1}^{r_2} \left\| \tilde{Z}_{k+j} - \rho(Z_{k+j-1}) \right\|_{L^2}^2 , \tag{18}$$

où $\tilde{Z}_{k+j}$ est la prévision, obtenue par (15), de la courbe Z_{k+j}.

4.1.5 Analyse des simulations

Deux simulations ont été réalisées en faisant varier les β_i.

- Simulation 1
  $\beta_1 = 0.8,\ \beta_2 = -0.9,\ \beta_3 = 0.5,\ \beta_4 = 0.6$ et $\beta_j = 0$ pour $j > 4$.
- Simulation 2
  $\beta_1 = 0.95,\ \beta_2 = 0.8,\ \beta_3 = -0.88,$ et $\beta_j = 0$ pour $j > 3$.

Pour chaque simulation, nous avons "laissé courir" le processus sur 100 trajectoires en 32 points de discrétisation. Les paramètres (dimension et lissage)

Méthodes d'approximation	**Critère (R)**	paramètre de lissage	dimension du s.e.p.
Interpolation linéaire	23.42	.	2
Lissage spline	22.48	0.01	2
Fonctions propres lisses	22.5	0.01	2
Lissage sous contrainte de rang	22.6	0.0004	2

TAB. 1 - *Minimum atteint par le critère (R) sur la première simulation et paramètres optimaux correspondants (lissage et dimension du sous-espace de projection).*

Méthode d'approximation	**E.Q.M.P.** $\times 100$
Interpolation linéaire	4.33
Lissage spline	4.10
Fonctions propres lisses	4.00
Lissage sous contrainte de rang	**3.86**

TAB. 2 - *Erreur de prévision des trajectoires du processus de la simulation* 1. *L'erreur de la méthode d'approximation la plus efficace est en caractères gras.*

ont été sélectionnés en utilisant les 80 premières trajectoires, $n = 80$, et les 4 suivantes, $r_1 = 4$, pour choisir ℓ et q. Les différentes techniques d'approximation ont ensuite été comparées sur les prévisions des 16 dernières trajectoires, $k = 84$ et $r_2 = 16$.

Simulation 1, rapport Signal/Bruit= 0.653 .

L'application du critère (R) nous a amené à choisir un sous-espace de dimension 2 pour représenter les trajectoires du processus alors que ρ est un opérateur de rang 4. Les autres dimensions ont sans doute été absorbées par le bruit. Il existe en effet une "cassure" entre la seconde valeur propre "théorique", $\mu_2 = 0.46$, et la troisième valeur propre, $\mu_3 = 0.03$, de l'opérateur de covariance et par conséquent une mauvaise estimation de la troisième "fonction" propre. Cela se traduit également dans la pratique par le choix d'un paramètre de lissage $\ell = 0.01$ relativement important. Les méthodes d'estimation avec lissage conduisent à de meilleurs prédicteurs que la méthode par interpolation linéaire au sens de l'E.Q.M.P. Le lissage sous contrainte de rang qui anticipe la réduction de dimension permet un lissage moins important ($\ell = 0.0004$) et autorise une meilleure prévision que lorsque lissage et réduction sont dissociés.

Simulation 2, rapport Signal/Bruit= 0.888 .

Toutes les méthodes conduisent à sélectionner un sous-espace de projection de dimension 2, la dernière ayant sans doute été abandonnée à cause de la faible valeur de la troisième valeur propre de l'opérateur de covariance du processus. L'approximation avec lissage sous contrainte de rang permet, selon notre critère, une prévision plus efficace.

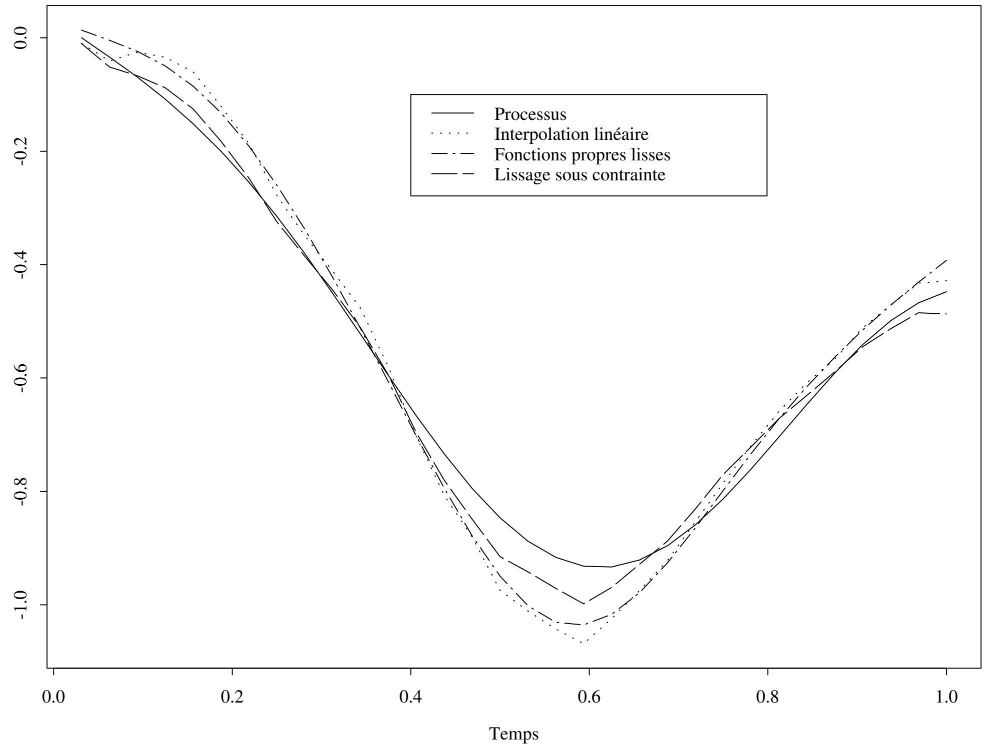

FIG. 2 - *Représentation, pour la deuxième simulation, des trajectoires prévues par le modèle avec lissage spline sous contrainte de rang, par la méthode d'estimation des fonctions propres lisses et par la méthode d'interpolation linéaire. La technique de lissage spline avant estimation des opérateurs a été omise pour ne pas surcharger le graphique car elle donne sensiblement la même prévision que les fonctions propres lisses : un "surlissage" de la prévision. La courbe en trait plein nommée "Processus" représente la valeur du processus privé de l'innovation.*

Méthodes d'approximation	Critère (R)	paramètre de lissage	dimension du s.e.p.
Interpolation linéaire	51.4	.	2
Lissage spline	50.9	0.0004	2
Fonctions propres lisses	51.4	6e-07	2
Lissage sous contrainte de rang	51.4	1e-07	2

TAB. 3 - *Choix des paramètres pour la deuxième simulation.*

Méthode d'approximation	E.Q.M.P. ×100
Interpolation linéaire	3.56
Lissage spline	3.75
Fonctions propres lisses	3.38
Lissage sous contrainte de rang	**2.62**

TAB. 4 - *Comparaison des erreurs quadratiques moyennes de prévision pour la deuxième simulation. L'erreur de la méthode d'approximation la plus efficace est en caractères gras.*

Nous avons constaté sur d'autres simulations qu'un rapport "Signal/Bruit" faible entraîne un lissage trop important de la partie prévisible des trajectoires et une sous-estimation de la dimension du sous-espace de représentation.

4.2 À partir d'une équation différentielle stochastique

Considérons un processus réel $(X_t;\ t \in \mathbb{R}^+)$ vérifiant l'équation différentielle stochastique (E.D.S.)

$$dX_t^{(q)} + \sum_{i=1}^{q-1} a_i dX_t^{(i)} = \sigma dW_t \tag{19}$$

où W_t est un mouvement brownien et $dX_t^{(i)}$ est la $i$ème dérivée du processus X_t, la $q$ème étant prise au sens de Itô. Les trajectoires de ce processus appartiennent à C_{q-1}. On suppose, pour des raisons de stationnarité, que les racines du polynôme $P(x) = \sum_{j=0}^{q} a_j x^j$ ont des parties réelles négatives.

Sous ces hypothèses, Bosq (1992) a démontré que le processus (X_t) pouvait se modéliser sous la forme d'un processus autorégressif hilbertien dont l'opérateur d'autocorrelation ρ est de rang fini q.

En effet,

$$E(X_{n+t} \mid X_s,\ s \leq n) = \sum_{j=0}^{q-1} X_n^{(j)} \phi_j(t),\ t \in [0,1],$$

où les ϕ_j sont des fonctions appartenant à C_{q-1} solutions de l'équation différentielle homogène ordinaire associée à (19).

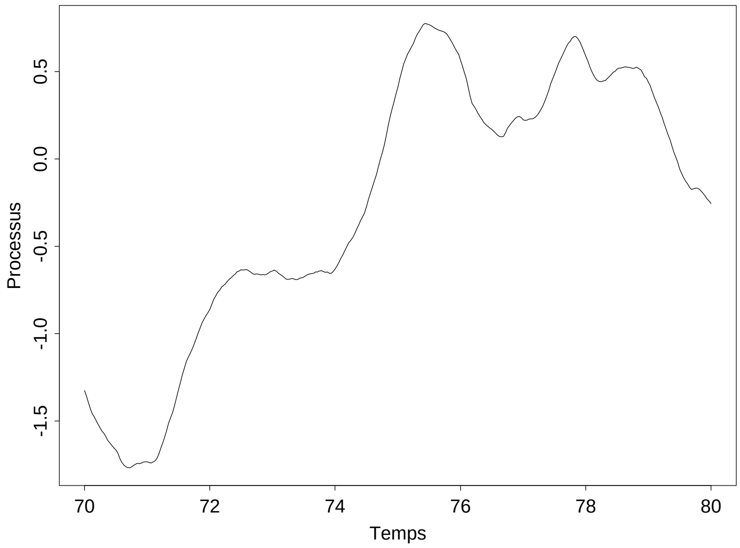

FIG. 3 - *Représentation d'une trajectoire simulée d'un processus ARH(1) gouverné par une E.D.S .*

4.2.1 Principe de la simulation

Plusieurs méthodes existent pour simuler de tels processus. Une manière élégante et exacte aurait été d'utiliser le résultat de Bergstrom (1983). Celui-ci montre que le processus discret obtenu à partir de la discrétisation en des points équidistants d'un processus vérifiant l'équation (19) suit une évolution de type ARMA$(q, q-1)$. Il définit de manière explicite les coefficients du processus ARMA en fonction des a_i et il lui suffit donc d'estimer les paramètres du modèle ARMA pour remonter à la structure continue. L'inconvénient majeur de sa méthode est que les coefficients du modèle ARMA sont compliqués à calculer. Nous avons donc dû abandonner ce schéma exact pour une approximation d'ordre 1 par le schéma de Milstein dont on peut trouver les fondements dans Kloeden & Platen (1992).

Ce schéma est le suivant :

$$\begin{cases} \dot{X}^{\delta}_{t+\delta} & = \dot{X}^{\delta}_{t} - (a_2 \dot{X}^{\delta}_{t} + a_1 X^{\delta}_{t})\delta + \sigma \delta W_t \\ X^{\delta}_{t+\delta} & = X^{\delta}_{t} + \delta \dot{X}^{\delta}_{t} \end{cases} \tag{20}$$

où $t = k\delta,\ k \in \mathbb{N}$ et $\delta W_t = W_{t+\delta} - W_t$.

4.2.2 Analyse des simulations

Nous avons simulé un processus $AR(2)$ à temps continu sur $[0, 80]$ en $p = 32$ points de discrétisations pour chaque unité de temps. Les coefficients sont $a_1 = 0.4$ et $a_2 = 0.8$. Les paramètres des modèles d'approximation spline ont été sélectionnés en fixant $r_1 = 7$ et $n = 60$ dans l'équation (17).

Méthodes d'approximation	**Critère (R)**	paramètre de lissage	dimension du s.e.p.
Interpolation linéaire	51.41	.	1
Lissage spline	51.39	0.01	1
Fonctions propres lisses	51.40	0.01	1
Lissage sous contrainte de rang	51.17	0.01	1

TAB. 5 - *Choix des paramètres pour la simulation d'une E.D.S.*

Méthode d'approximation	**E.Q.M.P.** $\times 100$
Interpolation linéaire	22.46
Lissage spline	22.45
Fonctions propres lisses	22.46
Lissage sous contrainte de rang	**22.41**

TAB. 6 - *Comparaison des erreurs quadratiques moyennes de prévision pour la simulation d'une E.D.S. L'erreur de la méthode d'approximation la plus efficace est en caractères gras.*

Les méthodes ont ensuite été comparées entre elles en fixant $r_2 = 14$ dans l'équation :

$$\text{E.Q.M.P.} = \frac{1}{r_2} \sum_{j=0}^{r_2-1} \frac{1}{p} \sum_{i=1}^{p} \left\{ \left[\widetilde{\mathbf{Z}}_{k+j} \right]_i - \left[\mathbf{Y}_{k+j} \right]_i \right\}^2 \tag{21}$$

où $\widetilde{\mathbf{Z}}_{k+j}$ est le vecteur contenant les prévisions du processus aux points de discrétisation sur le $(k+j)$ème intervalle et $\mathbf{Y}_{k+j}$ la trajectoire discrétisée observée. Ce critère de comparaison, qui est une approximation de l'erreur quadratique au sens de la norme $L^2[0,1]$, est moins précis que celui défini par (18), mais il ne nécessite pas de connaître l'opérateur d'autocorrélation ρ.

On constate que les différentes méthodes d'approximation fournissent des résultats "équivalents" à peu de choses près, le lissage sous contrainte de rang étant, une fois encore, meilleur que les autres.

À titre de comparaison, l'E.Q.M.P. ($\times 100$) réalisée par un modèle ARMA$(2,1)$ est de 10. Le fait que cette erreur soit nettement moins importante est logique car la discrétisation d'un processus AR(2) à temps continu engendre un processus ARMA$(2,1)$. On peut tout de même penser que la méthode non-paramétrique est concurentielle lorsque q est "grand".

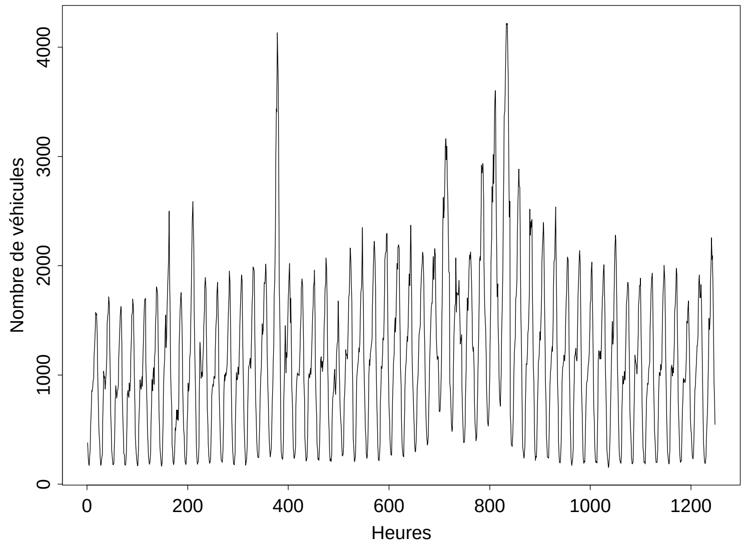

FIG. 4 - *Observation du trafic au péage de Saint-Arnoux le mercredi pendant l'année* 1990.

5 Comparaison sur des données réelles

5.1 Présentation des données

Ces données concernent le trafic autoroutier [2]. Pendant 1 an, il a été mesuré, heure par heure, le nombre de véhicules transitant par le péage de Saint-Arnoux, une sortie d'autoroute à proximité de Paris. Les données se présentent donc sous la forme d'un tableau de 365 colonnes et 24 lignes. À chaque colonne correspond un jour de l'année et à chaque ligne une heure.

Notre objectif est de prévoir le trafic d'une journée en fonction de celui du même jour de la semaine précédente. Pour cela, nous avons construit l'estimateur $\hat{\rho}$ de l'opérateur ρ à partir des observations du trafic sur les 30 premières semaines. Les 16 dernières semaines ont été utilisées pour mesurer la qualité des prévisions selon les méthodes employées.

5.2 Choix des paramètres

La dimension du sous-espace de représentation et le paramètre de lissage de nos estimateurs ont été sélectionnés à partir du critère de validation-croisée (R) défini par l'équation (17). Les 26 premières trajectoires ($n = 26$) ont été utilisées pour estimer les opérateurs de covariance et les 4 suivantes ($r_1 = 4$) pour choisir ℓ et q.

On constate que les méthodes d'approximation avec lissage permettent une réduction du rang de l'approximation de l'opérateur ρ par rapport à l'interpo-

2. Nous remercions le Département de génie informatique de l'Université technologique de Compiègne et la société Cofiroute pour nous avoir autorisés à exploiter ces données.

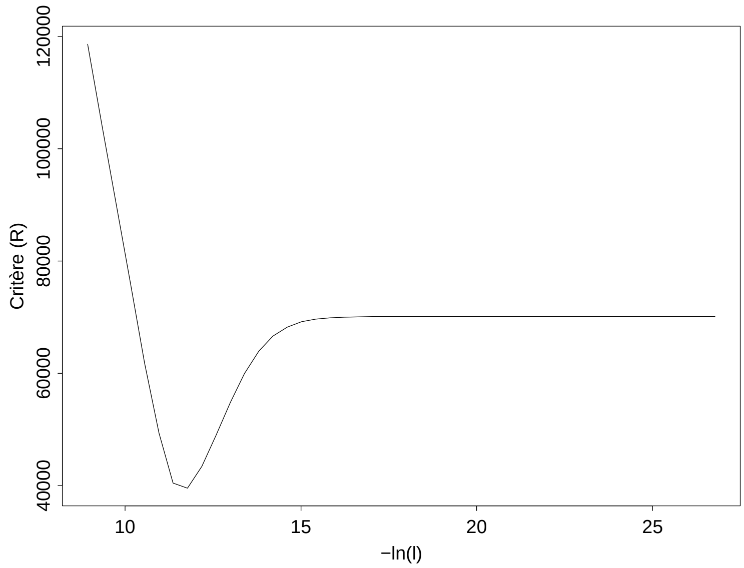

FIG. 5 - *Choix du paramètre de lissage, pour les données du trafic au péage de Saint-Arnoux, à l'aide du critère (R) pour le modèle d'estimation des trajectoires sous contrainte de rang* ($q = 1$). *Le minimum est atteint pour* $\ell = 7 \times 10^{-06}$.

Méthode d'approximation	**Critère (R)**	paramètre de lissage	dimension du s.e.p.
Interpolation linéaire	54000	.	3
Lissage spline	50800	0.0003	2
Fonctions propres lisses	51200	0.0004	2
Lissage sous contrainte de rang	40400	7e-06	1

TAB. 7 - *Choix des paramètres pour les données du trafic à Saint-Arnoux.*

lation linéaire.

5.3 Comparaison avec un modèle SARIMA

5.3.1 Construction du modèle

Bien que d'un point de vue théorique les modèles fonctionnels, qui opèrent sur des espaces fonctionnels, ne soient pas directement comparables avec les modèles de type SARIMA qui permettent d'effectuer des prévisions "ponctuelles", nous avons comparé les prévisions aux niveaux des points de discrétisation du processus.
Le problème posé par la non stationnarité du processus étudié, liée à la saisonnalité journalière et au trafic plus important les jours fériés, est résolu en considérant le modèle additif suivant:

$$X_t = \begin{cases} m_1(t) + Z_t & \text{si c'est un jour normal,} \\ m_2(t) + Z_t & \text{si c'est un jour férié .} \end{cases}$$

Ici, $m_1(t)$ et $m_2(t)$ sont deux vecteurs de taille $p = 24$ représentant la moyenne du processus le jour concerné tandis qu'un modèle de type SARIMA décrit Z_t. En notant B l'opérateur retard, $BZ_t = Z_{t-1}$, on a :

$$Y_t = (1 - B^{24})Z_t.$$

Le nouveau processus Y_t désaisonnalisé est supposé stationnaire. En adoptant la démarche de Box et Jenkins, tracé du corrélogramme et des autocorrélations partielles, nous avons choisi le modèle $(1, 0, 0) \times (0, 1, 1)^{24}$, soit encore :

$$(1 - \theta_1 B).(1 - B^{24})Z_t = (1 - \theta_2 B^{24})\epsilon_t,$$

où ϵ est un bruit blanc de variance σ^2.

Avec des mesures sur plus d'années consécutives, il aurait été possible de construire un modèle plus évolué qui tienne mieux compte des caractéristiques des jours fériés et des effets saisonniers comme, par exemple, dans le modèle de prévision de la demande d'électricité d'E.D.F. par Ernoult & Meslier (1982).

5.3.2 Estimation des paramètres

Les paramètres à estimer sont donc :

1. θ_1 et θ_2 ;

2. les effets des jours normaux et des jours de vacances $m_1(t)$ et $m_2(t)$;

3. la variance du bruit blanc.

Les paramètres du modèle SARIMA ont été estimés par maximum de vraisemblance sous Splus (Becker et coll. 1988), en supposant que l'innovation suivait une loi normale. Ils ont été estimés, comme dans le cas fonctionnel, à partir des données des 30 premières semaines.

Le test du "porte-manteau", d'hypothèse H_0 d'indépendance des résidus d'estimation, nous a conduits à accepter ce modèle.

Coefficient	Estimation	Variance
θ_1	0.94	1.7e-04
θ_2	0.89	3.3e-04

TAB. 8 - *Estimation des coefficients pour le modèle SARIMA.*

Méthode	E.Q.M.P.
Interpolation linéaire	24680
Lissage spline	30650
Fonctions propres lisses	32590
Lissage sous contrainte de rang	**21200**
Modèle SARIMA	47000

TAB. 9 - *Calcul de l'erreur quadratique moyenne de prévision sur les* 17 *dernières trajectoires.*

5.4 Comparaison par l'erreur quadratique moyenne de prévision

Les méthodes de prévision ont été comparées à l'aide du critère "erreur quadratique moyenne de prévision" défini par l'équation (21), appliqué entre le 37ème et le 54ème mercredi de l'année.

1. les méthodes fonctionnelles sont nettement plus efficaces que le modèle SARIMA pour prévoir le trafic autoroutier;

2. les deux premières méthodes de lissage n'apportent pas d'amélioration, au sens du critère précédent, par rapport à l'estimation sans lissage. La troisième, qui anticipe la réduction de rang de représentation des trajectoires, permet pour sa part d'améliorer assez nettement la prévision.

Conclusion

Au sens de l'erreur quadratique moyenne de prévision, la méthode de lissage avec contrainte de rang se révèle la plus efficace aussi bien sur les simulations que sur la prévision du trafic autoroutier à Saint-Arnoux. De plus, les méthodes fonctionnelles nonparamétriques sont faciles à mettre en œuvre et prévoient mieux, sur l'exemple traité, que les modèles SARIMA utilisés habituellement pour ce type de données. En contrepartie, elles nécessitent un grand nombre d'observations.

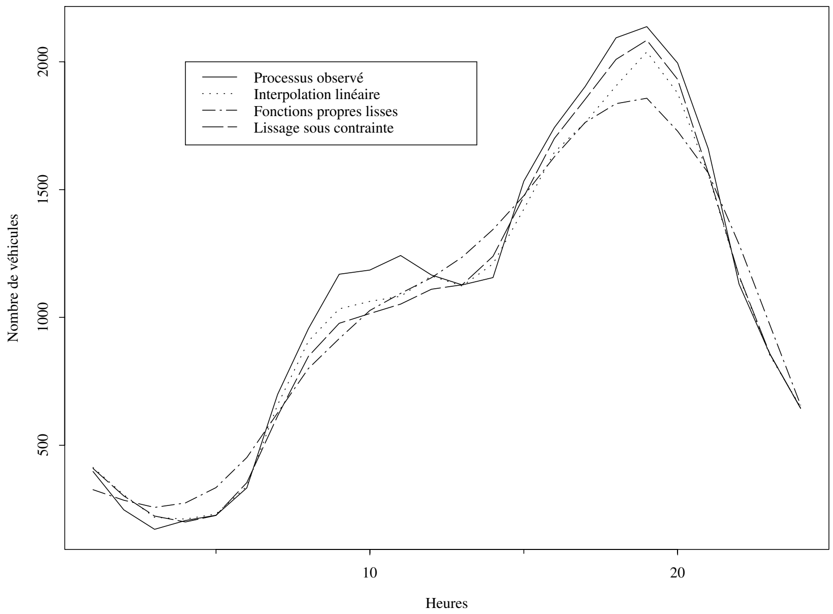

FIG. 6 - *Comparaison des méthodes fonctionnelles pour la prévision du trafic autoroutier d'un mercredi.*

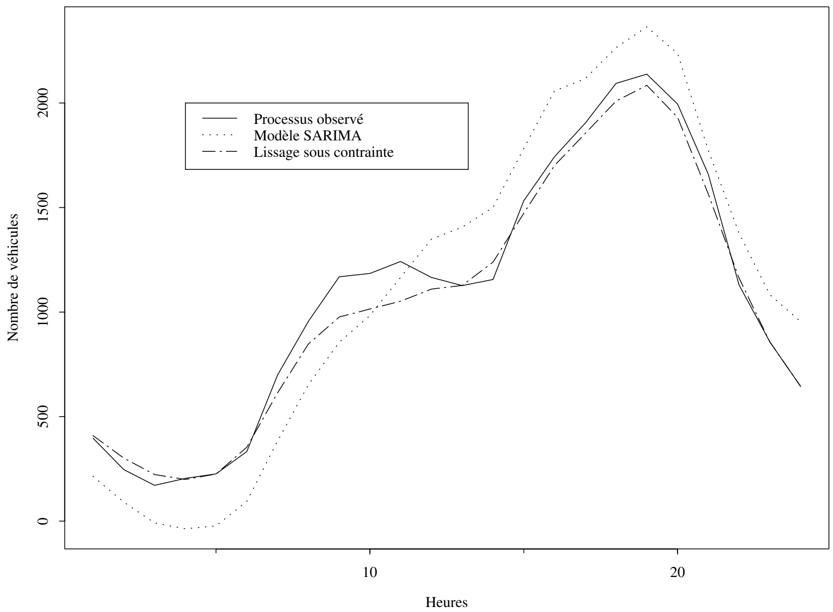

FIG. 7 - *Comparaison du modèle SARIMA et du modèle fonctionnel avec contrainte de rang pour la prevision du trafic d'un mercredi.*

References

Becker, R., Chambers, J., and Wilks, A. (1988). *The New S Language, a Programming Environment for Data Analysis and Graphics.* Wadsworth and Brooks/Cole.

Bergstrom, A. (1983). Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica*, 51(1):117–152.

Besse, P. (1991). Approximation spline de l'analyse en composantes principales d'une variable aléatoire hilbertienne. *Annales de la Faculté des sciences de Toulouse*, 12:329–346.

Besse, P., Cardot, H., and Ferraty, F. (1995). Simultaneous nonparametric regressions of unbalanced longitudinal data. *submitted for publication.*

Besse, P. and Pousse, A. (1992). Extension des analyses factorielles. In Droesbeke, J. et al., editors, *Modèles pour l'analyse des données multidimensionnelles*, pages 129–158. Economica.

Besse, P. and Ramsay, J. (1986). Principal component analysis of sampled curves. *Psychometrika*, 51:285–311.

Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. *Nato, Asi series.*

Bosq, D. (1992). Théorèmes limites pour les processus arb. *Publ. LSTA*, 158.

Bouhaddou, O., Obled, C., and Dinh, T. P. (1987). Principal component analysis and interpolation of stochastic processes: methods ans simulation. *Journal of Applied Statistics*, 14(3):251–267.

Box, G. and Jenkins, G. (1976). *Time Series Analysis.* Holden-Day.

Ernoult, M. et Meslier, F. (1982). Analyse et prévision de la demande d'energie électrique. *Revue générale d'électricité*, 4(82):257–277.

Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall.

Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, 46:1071–1085¸.

Kloeden, P. and Platen, E. (1992). *The Numerical Solution of Stochastic Differential Equations.* Springer-Verlag.

Pezzulli, S. and Silverman, B. (1993). On smoothed principal components analysis. *Computational Statistics*, 8:1–16.

Pumo, B. (1992). *Estimation et prévision de processus fonctionnels autorégressifs : Application aux processus à temps continu.* Thèse de troisième cycle, Université Paris 6.

Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society, B*, 53:539–572.

Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, B*, 53:233–243.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM.