Choix d'une covariance pour la prédiction par krigeage de séries chronologiques échantillonnées irrégulièrement

Emmanuel VAZQUEZ[1], Éric WALTER[2]

[1]Dpt. Signaux et Systèmes Électroniques
Supélec,
91192 Gif-sur-Yvette, France

[2]Laboratoire des Signaux et Systèmes
UMR CNRS – Supélec – Univ. Paris-Sud
91192 Gif-sur-Yvette, France
emmanuel.vazquez@supelec.fr,eric.walter@lss.supelec.fr

Résumé – Les processus aléatoires gaussiens à temps continu permettent de modéliser des séries chronologiques échantillonnées irrégulièrement. La prédiction des valeurs futures est un problème d'extrapolation par krigeage. Nous nous intéressons ici au choix de la covariance. Nous utilisons des combinaisons linéaires positives de covariances élémentaires pour améliorer la modélisation et les performances de prédiction des séries chronologiques échantillonnées irrégulièrement. Covariances sous forme de combinaisons linéaires positives de cosinus. Estimation des paramètres par méthodes REML et MAP.

Abstract – Continuous-time Gaussian processes make it possible to model irregularly sampled time series. The prediction of future values is a problem of extrapolation, which can in this context be addressed with the tools of Kriging. The issue considered here is the choice of the covariance to be used by the Kriging predictor. Positive linear combinations of elementary covariances are employed in order to improve modelling and predictive performance in complex cases where a simple classical covaraiance is not enough. The parameters of the resulting complex covariance are estimated by REML (restricted maximum likelihood) or MAP estimation. The procedure is applied to the well known lynx series , with very good results.

1 Introduction

Dans cet article, nous explorons le problème de la prédiction de séries chronologiques échantillonnées irrégulièrement en utilisant la méthode de prédiction appelée krigeage [2, 4, 5]. Nous traitons une série temporelle comme des observations échantillonnées d'une réalisation d'un processus gaussien à *temps continu*, ce qui permet de traiter indifféremment des échantillons observés régulièrement ou non. Les principes du krigeage sont brièvement rappelés dans la section 2. Afin d'effectuer une prédiction pertinente, il est nécessaire de choisir une covariance appropriée pour le processus aléatoire modélisant la série temporelle. Notre contribution consiste à proposer des covariances construites en assemblant un grand nombre de covariances élémentaires afin de décrire les caractéristiques des séries temporelles (section 3), ainsi qu'une méthode de type maximum a posteriori pour estimer les paramètres de ces covariances (section 4). La méthode sera appliquée dans la section 5 à un exemple typique du domaine des séries temporelles.

2 Prédiction par krigeage

2.1 Prédiction linéaire

Les principes de la prédiction linéaire de processus aléatoires (ou krigeage) sont très sommairement présentés dans cette section. Notons que ce type de prédiction peut être vue comme une régression régularisée [8]. Une série temporelle (supposée déterministe) est modélisée par un processus aléatoire stationnaire du second ordre noté $X(t)$, où $t \in \mathbb{R}$ est le paramètre temps, de moyenne inconnue et de covariance (à choisir)

$$k(h) = \mathrm{Cov}(X(t), X(t+h)), \quad t,h \in \mathbb{R}.$$

La série observée correspond donc à une réalisation $\boldsymbol{x}^{\mathrm{obs}} \in \mathbb{R}^n$ du vecteur aléatoire $(X(t_1),\ldots,X(t_n))^{\mathsf{T}}$. Si en outre les observations sont corrompues par un bruit de mesure additif, ce bruit est modélisé par des variables aléatoires indépendantes N_i. Pour simplifier la présentation, nous supposerons ici le bruit négligeable.

Supposons dans un premier temps la moyenne de $X(t)$ connue. La méthode la plus simple pour prédire $X(t)$ est de calculer la meilleure projection linéaire $\hat{X}(t)$ de $X(t)$ sur l'espace $\mathscr{H}_S$ généré par les variables aléatoires observées $X(t_i)$. Ceci correspond à la formulation du krigeage, et signifie que l'on cherche un prédicteur linéaire $\widehat{X}(t) = \sum_i \widehat{\lambda}_{i,t} X(t_i)$ tel que

$$\mathrm{Var}(\widehat{X}(t) - X(x)) = \|\widehat{X}(t) - X(x)\|^2$$

soit minimum ou, de manière équivalente puisque la meilleure prédiction linéaire est la projection orthogonale sur $\mathscr{H}_S$, tel que

$$\begin{aligned}(\widehat{X}(t) - X(t), X(t_i)) &= \mathrm{Cov}[\widehat{X}(t) - X(x), X(t_i)] \\ &= 0, \quad \forall i \in \{1,\ldots,n\}.\end{aligned} \tag{1}$$

En remplaçant $\widehat{X}(t)$ par son expression en fonction des $X(t_i)$, on vérifie que le vecteur $\widehat{\boldsymbol{\lambda}}_t = (\widehat{\lambda}_{1,t},\ldots,\widehat{\lambda}_{n,t})^{\mathsf{T}}$ est solution du

système linéaire

$$\boldsymbol{K}\widehat{\boldsymbol{\lambda}}_t = \boldsymbol{k}_t\,, \tag{2}$$

où $\boldsymbol{K}$ est la matrice des covariances $k(t_i - t_j)$ et $\boldsymbol{k}_t$ est le vecteur des covariances $k(t-t_i)$. La matrice $\boldsymbol{K}$ est de rang plein lorsque la covariance est définie positive et que les observations ne sont pas répétées. Notons que le prédicteur obtenu est sans biais, puisque la moyenne de $X(t)$ est connue.

2.2 Krigeage intrinsèque

Supposons maintenant que $\mathrm{E}[X(t)] = b$, avec $b \in \mathbb{R}$ inconnu. On remarque que les combinaisons linéaires telles que $\sum_i \lambda_i X(t_i)$ avec $\sum_i \lambda_i = 0$ filtrent la moyenne inconnue de $X(t)$ dans le sens où $\mathrm{E}[\sum_i \lambda_i X(t_i)] = 0$. L'idée est donc de se ramener au cas à moyenne nulle en utilisant de tels accroissements. On considère plus généralement des accroissements d'ordre l de $X(t)$ définis par les variables aléatoires

$$X(\lambda) = \sum_{i=1}^{n} \lambda_i X(t_i) \tag{3}$$

telles que pour tout $r \in \{0,\ldots,l\}$,

$$\sum_{i=1}^{n} \lambda_i t_i^r = 0\,. \tag{4}$$

Cette dernière propriété peut être vue comme l'orthogonalité de la mesure à support fini $\lambda = \sum_{i=1}^{n} \lambda_i \delta_{t_i}$ avec les monômes t^r (δ_t désigne la mesure de Dirac telle que pour tout $B \subset \mathbb{R}$, $\delta_t(B)$ est égal à un si $t \in B$ et zéro sinon). Soit $\tilde{\Lambda}_l$ l'ensemble des mesures à support fini vérifiant la relation (4). Les éléments de $\tilde{\Lambda}_l$ filtrent toute moyenne polynomiale de $X(t)$ (d'ordre inférieur ou égal à l).

Une *covariance généralisée* [5] notée $k(t,s)$ permet de calculer la covariance de $X(\lambda)$ et $X(\mu)$, $\lambda, \mu \in \tilde{\Lambda}_l$, en utilisant la relation

$$\mathrm{Cov}[X(\lambda), X(\mu)] = \sum_{i,j} \lambda_i k(t_i, t_j) \mu_j\,. \tag{5}$$

Les covariances généralisées sont de type *conditionnellement positif* [5], c'est-à-dire qu'elles doivent garantir

$$\mathrm{Var}[X(\lambda)] = \sum_{i,j} \lambda_i \lambda_j k(t_i, t_j) \geq 0$$

pour toute mesure $\lambda \in \tilde{\Lambda}_l$. La classe des covariances est donc incluse dans celle des covariances généralisées.

Le krigeage dit *intrinsèque* [5] construit un prédicteur $\widehat{X}(t) = \sum_{i=1}^{n} \widehat{\lambda}_{i,t} X(t_i)$ minimisant $\mathrm{Var}[\widehat{X}(t) - X(t)]$ sous la contrainte que $\widehat{X}(t) - X(t)$ soit un accroissement d'ordre l, c'est-à-dire $\sum_i \widehat{\lambda}_{i,t} \delta_{t_i} - \delta_t \in \tilde{\Lambda}_l$. On est ainsi amené a résoudre le système linéaire:

$$\begin{pmatrix} \boldsymbol{K} & \boldsymbol{P}^{\mathsf{T}} \\ \boldsymbol{P} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_t \\ \boldsymbol{\mu}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{k}_t \\ \boldsymbol{p}_t \end{pmatrix}, \tag{6}$$

où $\boldsymbol{P}$ est une matrice $(l+1) \times n$ dont les éléments sont les t_i^r, $\boldsymbol{\mu}_t$ est un vecteur de coefficients de Lagrange et $\boldsymbol{p}_t$ est le vecteur des monômes t^l.

La théorie du krigeage intrinsèque et des fonctions conditionnellement positives est mathématiquement plus difficile que celle du krigeage des processus aléatoires à moyenne connue. Sa mise en œuvre reste toutefois d'un niveau de simplicité élementaire et, outre le fait qu'il est possible de traiter les processus aléatoires à moyenne inconnue, le krigeage intrinsèque permet de considérer une classe de covariances plus vaste (notons, par exemple, que les splines de type "plaques minces" [10] sont fondées sur des noyaux conditionnellement positifs). Le krigeage intrinsèque permet également d'inclure de l'information a priori qu'on ne souhaite pas qu'elle soit régularisée (l'idée étant alors d'inclure des fonctions particulières en plus des monômes considérés plus haut [9]).

3 Choix d'une fonction de covariance

Nous cherchons à écrire la covariance de $X(t)$ sous la forme d'une combinaison linéaire *positive* de covariances élémentaires stationnaires. Comme cas particulier, nous pouvons considérer une somme de cosinus telle que

$$k_{\boldsymbol{\alpha}}(h) = \sum_{i=1}^{l} e^{\alpha_i} \cos u_i h\,, \tag{7}$$

où les pulsations u_i peuvent être choisies a priori et où les α_i sont les paramètres ajustables de cette covariance. Notons que le caractère périodique de $k(h)$ n'est pas gênant si l'on prend soin de choisir une période plus grande que l'horizon d'étude de $X(t)$.

4 Estimation des paramètres

Nous souhaitons estimer les paramètres α_i lorsque l est grand (typiquement quelques centaines). La forme paramétrique (7) s'apparente alors à une représentation dans le domaine de Fourier. Une première idée consiste à estimer les α_i par maximum de vraisemblance restreint (*REML*) [7]. Cette méthode est préférée à celle du maximum de vraisemblance parce que la moyenne de $X(t)$ est inconnue.

4.1 Maximum de vraisemblance restreint

L'estimation au sens du maximum de vraisemblance restreint (*restricted maximum likelihood*, abrégé par *REML*, en anglais) des paramètres de la covariance de $X(t)$ consiste à écrire non pas la fonction de vraisemblance des données observées, mais celle des accroissements (ou accroissements généralisés) de ces données. Ces accroissements s'appellent aussi des *contrastes*.

Notons $\boldsymbol{X}^{\mathrm{obs}} = (X(t_1), \ldots, X(t_n))^{\mathsf{T}}$ le vecteur aléatoire des observations (en supposant le bruit négligeable pour simplifier). Notons également $\boldsymbol{P} = (t_j^i)_{i=0,j=1}^{l,n}$ la matrice $(l+1) \times n$ des monômes de degré inférieur ou égal à l évalués sur $S = \{t_1, \ldots, t_n\}$. L'espace des mesures à support S annulant les polynômes de degré inférieur ou égal à l est de dimension $n - l - 1$. Supposons trouvée une matrice $\boldsymbol{W}$ de taille $n \times (n-l-1)$ et de rang $n-l-1$, telle que

$$\boldsymbol{P}\boldsymbol{W} = \boldsymbol{0}\,.$$

(Les colonnes de $\boldsymbol{W}$ sont dans le noyau de $\boldsymbol{P}$.) Notons que les colonnes de $\boldsymbol{W}$ sont donc les coefficients de mesures à support S, $\sum_{j=1}^{n} \boldsymbol{W}_{[i,j]} \delta_{t_j} \in \tilde{\Lambda}_l$. Alors $\boldsymbol{Z} = \boldsymbol{W}^{\mathsf{T}} \boldsymbol{X}^{\mathrm{obs}}$ est un vecteur aléatoire gaussien à valeurs dans $\mathbb{R}^{n-l-1}$, de moyenne nulle et de matrice de covariance $\boldsymbol{W}^{\mathsf{T}} \boldsymbol{K}(\boldsymbol{\alpha}) \boldsymbol{W}$, où $\boldsymbol{K}(\boldsymbol{\alpha})$ est la matrice symétrique de covariance généralisée ayant comme éléments les scalaires $k_{\boldsymbol{\alpha}}(t_i - t_j)$. Le vecteur aléatoire $\boldsymbol{Z}$ est

un *vecteur de contrastes*. La log–vraisemblance des contrastes s'écrit

$$L(\boldsymbol{z},\boldsymbol{\alpha}) = -\frac{n-l-1}{2}\log 2\pi - \frac{1}{2}\log\det(\boldsymbol{W}^\mathsf{T}\boldsymbol{K}(\boldsymbol{\alpha})\boldsymbol{W}) - \frac{1}{2}\boldsymbol{z}^\mathsf{T}(\boldsymbol{W}^\mathsf{T}\boldsymbol{K}(\boldsymbol{\alpha})\boldsymbol{W})^{-1}\boldsymbol{z}. \quad (8)$$

Plusieurs méthodes peuvent être envisagées pour calculer la matrice $\boldsymbol{W}$. Les approches proposées par [3, 6, 7] sont considérées comme classiques. Nous préférons utiliser la décomposition *QR* de $\boldsymbol{P}^\mathsf{T}$

$$\boldsymbol{P}^\mathsf{T} = (\boldsymbol{Q}_1 \mid \boldsymbol{Q}_2)\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix},$$

où $(\boldsymbol{Q}_1 \mid \boldsymbol{Q}_2)$ est une matrice orthogonale de taille $n\times n$ et $\boldsymbol{R}$ une matrice triangulaire supérieure de taille $(l+1)\times(l+1)$. Il est immédiat de vérifier que les colonnes de la matrice $\boldsymbol{Q}_2$ forment une base du noyau de $\boldsymbol{P}$ et nous pouvons donc choisir $\boldsymbol{W} = \boldsymbol{Q}_2$. Notons que $\boldsymbol{W}^\mathsf{T}\boldsymbol{W} = \boldsymbol{I}_{n-l-1}$. Le coût algorithmique du calcul de la vraisemblance est en $O(n^3)$.

4.2 Régularisation sur les paramètres

Comme nous l'illustrerons dans la section 5, l'estimation au sens du maximum de vraisemblance permet d'obtenir une covariance $k_{\boldsymbol{\alpha}}$ très proche de la covariance empirique. Cependant, s'il existe une bonne fidélité entre le modèle de covariance et la covariance empirique, la capacité de prédiction (ou de généralisation) du modèle aléatoire n'est pas nécessairement satisfaisante. Il s'agit d'un phénomène de sur-adaptation aux données qui s'explique par le fait que le modèle comporte un grand nombre de paramètres.

Il apparaît alors nécessaire de régulariser les paramètres α_i, ce qui peut se faire par estimation au sens du maximum a posteriori (MAP) avec un a priori de régularité sur les paramètres. Comme ces paramètres s'apparentent à une estimée de la densité spectrale et que nous voulons que cette estimée soit relativement régulière, l'idée proposée est de modéliser les paramètres α_i par un processus gaussien $\alpha(u)$, tel que $\alpha_i = \alpha(u_i)$, $i = 1,\ldots,l$. Autrement dit, la densité de probabilité du vecteur aléatoire $\boldsymbol{\alpha} = (\alpha(u_1),\ldots,\alpha(u_l))^\mathsf{T}$ s'écrit sous la forme

$$p(\boldsymbol{\alpha}) = \frac{1}{Z}\exp\left(-\frac{1}{2}(\boldsymbol{\alpha}-\bar{\boldsymbol{\alpha}})^\mathsf{T}\boldsymbol{K}_{\boldsymbol{\alpha},\mathrm{reg}}^{-1}(\boldsymbol{\alpha}-\bar{\boldsymbol{\alpha}})\right),$$

où Z est une constante de normalisation, $\bar{\boldsymbol{\alpha}}$ est la moyenne du vecteur $\boldsymbol{\alpha}$, et $\boldsymbol{K}_{\boldsymbol{\alpha},\mathrm{reg}}$ est sa matrice de covariance. Nous faisons l'hypothèse que la moyenne du processus $\alpha(u)$ est inconnue mais possède une forme linéairement paramétrée, par exemple du type $b_0 + b_1 u$ ou une autre forme polynomiale. Reste à effectuer le choix de la covariance de $\alpha(u)$. Nous avons choisi une covariance de type exponentiel [11] en ajustant ses paramètres par une validation croisée rudimentaire.

Avec cet a priori, nous estimons $\boldsymbol{\alpha}$ au sens du MAP, en maximisant

$$J(\boldsymbol{\alpha}) = -\frac{1}{2}\log\det(\boldsymbol{W}^\mathsf{T}\boldsymbol{K}(\boldsymbol{\alpha})\boldsymbol{W}) - \frac{1}{2}\boldsymbol{z}^\mathsf{T}(\boldsymbol{W}^\mathsf{T}\boldsymbol{K}(\boldsymbol{\alpha})\boldsymbol{W})^{-1}\boldsymbol{z} - \frac{1}{2}\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{W}_{\boldsymbol{\alpha}}(\boldsymbol{W}_{\boldsymbol{\alpha}}^\mathsf{T}\boldsymbol{K}_{\boldsymbol{\alpha},\mathrm{reg}}\boldsymbol{W}_{\boldsymbol{\alpha}})^{-1}\boldsymbol{W}_{\boldsymbol{\alpha}}^\mathsf{T}\boldsymbol{\alpha} \quad (9)$$

par rapport à $\boldsymbol{\alpha}$. L'expression (9) est constituée de deux parties. La première correspond à la log-vraisemblance restreinte formée à partir d'une matrice de contrastes $\boldsymbol{W}$ de taille $n\times(n-q)$, où q est la dimension de l'espace $\mathscr{N}$ des fonctions polynomiales contenant la moyenne inconnue de $X(t)$. Le vecteur des contrastes $\boldsymbol{z}$ s'obtient donc par la transformation linéaire $\boldsymbol{z} = \boldsymbol{W}^\mathsf{T}\boldsymbol{x}^{\mathrm{obs}}$. La matrice $\boldsymbol{K}(\boldsymbol{\alpha})$ désigne la matrice de covariance du vecteur des observations, formée à partir de $k_{\boldsymbol{\alpha}}(h)$ paramétrée par $\boldsymbol{\alpha}$. Le dernier terme de (9) correspond à la log-densité restreinte du vecteur des paramètres formée à partir d'une matrice de contrastes $\boldsymbol{W}_{\boldsymbol{\alpha}}$. Nous utilisons cette formulation pour ne pas avoir à prendre en compte la moyenne inconnue du processus $\alpha(u)$.

Il est aisé de maximiser (9) par rapport aux paramètres α_i parce que le gradient possède une expression analytique simple.

5 Exemple

Pour illustrer la méthode, considérons la série chronologique classique du nombre de lynx au Canada entre 1821 et 1934 représentée sur la figure 1. Cette série est étudiée en détails dans [1], ce qui permet de comparer facilement les performances des modèles proposés. Nous la modélisons par un processus gaussien stationnaire $X(t)$, $t\in\mathbb{R}$, de moyenne inconnue et de covariance $k(h)$ à choisir. La série observée correspond donc à une réalisation $\boldsymbol{x}^{\mathrm{obs}}\in\mathbb{R}^n$ du vecteur aléatoire $(X(t_1),\ldots,X(t_n))^\mathsf{T}$ et la prédiction est un problème d'extrapolation par krigeage. La difficulté du problème est de choisir une covariance appropriée pour $X(t)$.

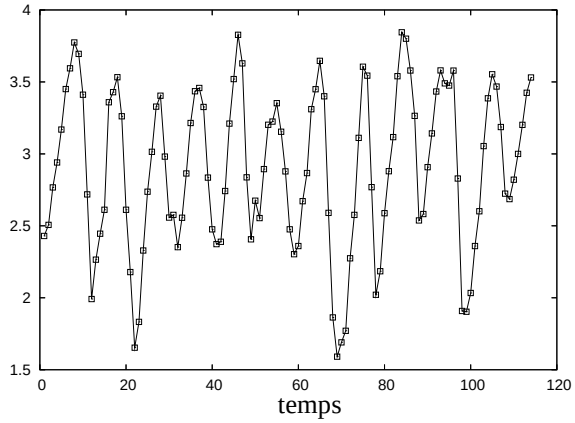

Figure 1: Évolution du logarithme en base 10 du nombre de lynx au Canada sur la période 1821–1934 (les abscisses correspondent à des pas de temps en années).

Le caractère périodique de la série explique la forme oscillante de sa fonction d'autocovariance empirique, présentée à la figure 2. Ceci suggère de prendre une covariance $k(h)$ avec des oscillations. Toutefois, des expériences numériques avec des modèles de covariance simples (c'est-à-dire comportant relativement peu de paramètres) ne conduisent pas à des résultats satisfaisants[1]. Notons que le modèle proposé par [1] est un *AR* d'ordre 12 (sélectionné d'après un critère d'Akaike), ce qui signifie qu'il a typiquement six modes de résonance. Le modèle de covariance de cet $AR(12)$ possède donc une capacité d'adaptation aux données plus importante que les covari-

[1] Ces expériences numériques sont menées avec des covariances de type cosinus amorti.

ances à temps continu mentionnées ci-dessus. Ceci souligne la faiblesse des procédures classiques de choix de covariance où l'on se limite à choisir celles-ci dans des familles restreintes [2, 7, 11].

D'après la figure 2, la covariance à temps continu obtenue par estimation *REML* apparaît qualitativement proche de la covariance empirique. Pour évaluer la qualité du modèle, nous utilisons le même critère S que dans [1], c'est-à-dire la moyenne quadratique des erreurs de prédictions à un pas sur les 14 dernières valeurs de la série. S vaut 0.138 dans le cas de l'$AR(12)$ proposé par [1] et environ 0.350 dans notre cas. Notre résultat de prédiction à un pas après estimation de la covariance par maximum de vraisemblance est donc très mauvais en raison d'un phénomène de sur-adaptation aux données qui s'explique par le fait que le modèle de covariance comporte plus de paramètres qu'il n'y a de données.

Il apparaît alors nécessaire de régulariser les paramètres α_i et nous utilisons la méthode proposée dans la section 4.2. Sur la série des lynx, nous obtenons alors des valeurs S entre 0.120 et 0.125, selon le choix des paramètres de la covariance de $\alpha(u)$. Les performances de prédiction sont donc *meilleures* que celles obtenues avec le modèle $AR(12)$ de [1]. La covariance estimée est présentée à la figure 2. Les paramètres $\hat{\alpha}_i$ estimés sont représentés en fonction des pulsations u_i sur la figure 3 où nous avons également représenté la densité spectrale du processus $AR(12)$ de [1]. On constate la ressemblance entre les deux modèles. L'avantage du modèle proposé est que l'on contrôle de manière flexible l'adaptation aux données.

En conclusion, la méthode proposée, qui s'applique aussi à des séries non uniformément échantillonnées, se révèle pertinente et pourrait constituer la base de futurs travaux.

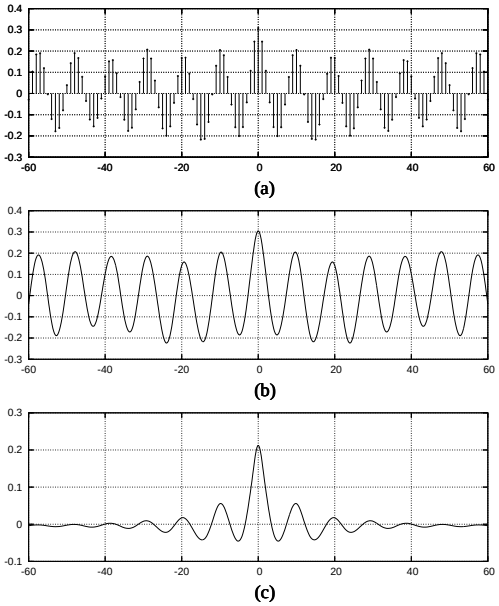

Figure 2: Fonctions de covariance de la série des lynx; (a) covariance à temps discret estimée empiriquement (estimateur *non biaisé*); (b) covariance à temps continu estimée par maximum de vraisemblance; (c) covariance à temps continu estimée par maximum a posteriori.

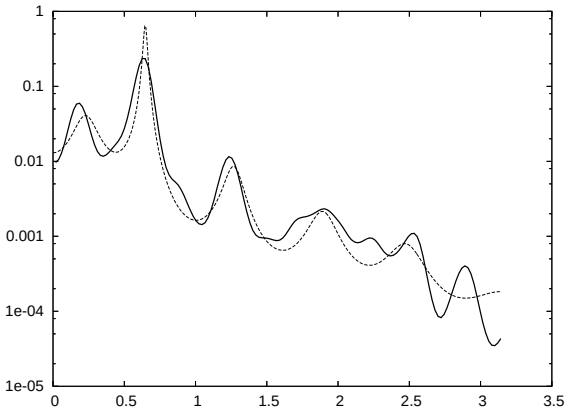

Figure 3: Représentation des paramètres estimés (et renormalisés) de la covariance en fonction de la pulsation (trait continu). Densité spectrale du modèle $AR(12)$ de [1] (trait interrompu).

References

[1] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 1987.

[2] J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 1999.

[3] D. A. Harville. Bayesian inference for variance components using only the error contrasts. *Biometrika*, 61:383–385, 1974.

[4] G. Matheron. Principles of geostatistics. *Economic Geology*, 58:1246–1266, 1963.

[5] G. Matheron. The intrinsic random functions, and their applications. *Adv. Appl. Prob.*, 5:439–468, 1973.

[6] H. D. Patterson and R. Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554, 1971.

[7] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

[8] E. Vazquez. *Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications*. PhD thesis, Université Paris XI Orsay, 2005.

[9] E. Vazquez and E. Walter. Intrinsic kriging and prior information. *Applied Stochastic Models in Business and Industry*, 21(2):215–226, 2005.

[10] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1990.

[11] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. Springer Series in Statistics. Springer-Verlag, New york, 1986.