
PUBLICATIONS DU GROUPE DE TRAVAIL STAPH:
STATISTIQUE FONCTIONNELLE ET OPÉRATORIELLE

STAPH-2012-01 *Recueil de résumés de l'année 2011-2012*

ALAIN BOUDOU, FRÉDÉRIC FERRATY, YVES ROMAIN, PASCAL SARDA, PHILIPPE VIEU
ET SYLVIE VIGUIER-PLA

Institut de Mathématiques, Université Paul Sabatier, Toulouse, France

STAPH: Groupe de travail en Statistique Fonctionnelle et Opératoireielle. Présentation des activités 2011-2012.

**Alain BOUDOU, Frédéric FERRATY, Yves ROMAIN, Pascal SARDA,
Philippe VIEU et Sylvie VIGUIER-PLA**

Adresse pour correspondance:
Institut de Mathématiques, Université Paul Sabatier
Toulouse

Dans ce document, sont regroupés les résumés des exposés donnés lors des séances du groupe de travail en Statistique Fonctionnelle et Opératoireielle STAPH au cours de l'année 2011-2012. Nous tenons tout d'abord à remercier chaleureusement tous les orateurs pour la qualité de leurs contributions qui ont donné lieu à des échanges fructueux. Nous saluons tout particulièrement les nouveaux venus à notre groupe de travail (la moitié des orateurs) et avec lesquels des membres de STAPH ont noué des liens scientifiques ou amorcé des collaborations. De manière générale, nous avons accueilli des doctorants, des jeunes chercheurs ou encore des chercheurs confirmés, suivant en cela un objectif de pluralité qui nous tient à coeur.

Cette introduction est également l'occasion pour nous de rappeler qu'un des points forts de l'activité de notre groupe de travail lors de cette année 2012 a été son implication dans l'organisation de plusieurs sessions invitées lors de conférences thématiquement diversifiées, comme par exemple la conférence mondiale de Statistique Nonparamétrique (voir [4]) et la conférence européenne annuelle d'Econométrie organisée par le groupe européen ERCIM (voir [1]). Il faut aussi souligner que, comme d'habitude, nous avons pris une part importante dans la préparation des journées nationales de Statistique Fonctionnelle qui se sont déroulées cette année à Montpellier (voir [6]). Ces journées nationales n'avaient naturellement pas eu lieu en 2011, année pour laquelle nous nous étions concentrés sur l'organisation du second meeting international en Statistique Fonctionnelle et Opératoireielle (IWFOS2011) qui s'était tenu en Juin à Santander en Espagne sous la co-présidence franco-espagnole de Juan Cuesta-Albertos, Frédéric Ferraty (Toulouse) et Wenceslao Gonzalez-Manteiga (rappelons que l'ouvrage [2] synthétise les contributions à cette conférence). Rappelons aussi la parution récente du manuel de base en Statistique Fonctionnelle (voir [3]) qui est directement issu de la première édition de la conférence IWFOS2008 (qui avait eu lieu à Toulouse) et

auquel ont contribué les plus grands spécialistes mondiaux par le biais de différents chapitres d'ordre méthodologique et/ou bibliographique.

Pour terminer, il faut ajouter que le succès retentissant des deux premières éditions de la conférence internationale IWFOs nous a décidé à reconduire une troisième fois l'expérience pour 2014. Ainsi nous sommes fiers et heureux de vous annoncer que la troisième édition de IWFOs sera le fruit d'une co-organisation franco-italienne et qu'elle aura lieu en 2014. Les toutes premières informations sont d'ores et déjà disponibles (voir [5]).

Références

- [1] Aneiros, G. and Vieu, P. (Org.) Session on Recent Advances in Functional Data Analysis. ERCIM conference Oviedo, Spain, December 2012 <http://www.cfe-csda.org/ercim12/>.
- [2] Ferraty, F. (Ed). Recent advances in functional data analysis and related topics, Springer, Contributions to Statistics, 2011.
- [3] Ferraty, F. and Vieu, P. (Org.) Session on Infinite Dimensional Statistics. ISNPS conference June 15-19, 2012. <http://www.isnpstat.org/>.
- [4] Ferraty, F. and Romain, Y. (Eds). The Oxford handbook on functional data analysis, Oxford University Press, 2011.
- [5] Goia, A. and Vieu, P. (Org.). International Workshop on Functional and Operatorial Statistics. June 19-21, 2014. <http://semeq.unipmn.it/iwfos2014/>.
- [6] Mas, A. (Org.) 7èmes journées de Statistique Fonctionnelle. <http://www1.montpellier.inra.fr/mistea/jsf2012/index.htm>.

Multiple Change Point Detection

Jaromir ANTOCH*, **Daniela JARUŠKOVÁ**

* Adresse pour correspondance:
Department of Probability and Mathematical Statistics
Faculty of Mathematics and Physics
Charles University of Prague
Czech Republic
e-mail : jaromir.antoch@mff.cuni.cz

Abstract

Detection of multiple changes and/or data segmentation are among basic problems we encounter in statistics and data analysis. In this paper we concentrate on multiple change-point detection in the mean of a series of independent random variables. We suggest to apply a maximum type test statistic. Our primary focus is on the effective calculation of critical values for (very) large sample sizes and moderate or large number of segments. To that purpose, a combination of Monte Carlo simulations and Bellman's principle of optimality is used. It is shown that, indisputably, computer memory is a critical bottleneck in solving the considered problem. Thus, minimization of the memory requirements and appropriate order of calculations appear to be the keys to success. In addition, we also thoroughly discuss computational issues related to obtaining approximate critical values using the theory of exceedance probability of Gaussian fields over a high level.

Application des méthodes d'analyses des données pour la gestion des ressources naturelles

Souffien MOSBAH^{ab}, Mohamed GHORBEL^a, Othman JARBOUI^{ab}, Philippe LENFANT^b, Marion VERDOIT- JARRAYA^a

Adresses pour correspondance:

^a : CEntre de Formation et de Recherche sur les Environnements Méditerranéens (CEFREM), Perpignan France.

^b : Institut National des Sciences et Technologies de la Mer (INSTM), Sfax Tunisie.
e-mail: souffien2004@yahoo.fr

1 Introduction

La Tunisie, avec 1250 km de côtes, assure une production halieutique moyenne annuelle estimée à environ 100000 tonnes et occupe la seconde place dans l'exportation avec une valeur estimée à 126 Millions d'Euros/an. Le golfe de Gabès, s'étendant sur 750 km de côtes, se situe vers la partie méridionale. Son cadre géomorphologique exceptionnel en Méditerranée, lui a offert une richesse biologique et une biodiversité assez importante et il fut souvent qualifié de nurserie et de pépinière de la Méditerranée (Hattour, 1991). Cette richesse a été à l'origine d'un développement des activités de pêche et l'intense effort de pêche déployé dans cette région a causé l'apparition de signes de sur-exploitation de certaines espèces : le pageot *Pagellus erythrinus* (Ghorbel, 1996), la seiche *Sepia officinalis* (Ben Meriem et al., 1994a) et le rouget rouge *Mullus surmuletus* (Ben Meriem et al., 1994b). Le golfe de Gabès est une pêcherie multispécifique et pluriengins (Missaoui et al., 2000 ; Ben Meriem et al., 2005) ; deux activités prédominant dans cette région : la pêche côtière et la pêche au chalutage benthique. Malgré leur importance, rares sont les études qui ont été dédiées à ces activités (Jabeur et al., 2000 ; Missaoui et al., 2000).

L'aménagement du secteur de la pêche requiert des connaissances sur les ressources exploitables, sur l'état des stocks mais aussi des informations sur les pêcheries actives et le contexte socio-économique des différentes activités. La détermination des variables influençant la production s'avère d'une importance capitale. Cette variation peut être liée aux facteurs

spatio-temporels (Bertignac, 1992), environnementaux (Allen and Punsly, 1984; Bigelow et al., 2002) et à l'effort de pêche déployé (Hilborn and Walters, 1992). En Tunisie, l'essentiel des études se sont focalisées sur l'aspect biologique et dynamique des espèces exploitées, négligeant à cet effet l'aspect halieutique et socio-économique des activités. La présente étude est une approche écosystémique basée essentiellement sur l'utilisation de méthodes d'analyses multivariées qui visent à caractériser les principales activités d'un point de vue halieutique et économique et suivre leur évolution spatio-temporelle.

2 Matériel et Méthodes

2.1 Données

La base de données provient essentiellement de séries d'enquêtes personnelles effectuées dans le plus grand port de la Tunisie dans la ville de Sfax. L'échantillonnage est aléatoire simple. Un questionnaire comportant des informations sur les caractéristiques techniques des unités, la production par espèce et les zones exploitées a été rempli par les commandants de bord. Une autre base de données décennale a été fournie par les services statistiques de la Direction Régionale de la Pêche de Sfax ; elle contenait des informations sur les quantités d'espèces capturées par mois, par type de pêche et par port de débarquements.

2.2 Méthodes

Cette étude est basée essentiellement sur l'utilisation de méthodes statistiques d'analyses multivariées des données. Ces méthodes sont de plus en plus employées en halieutique (Pelletier et Ferraris, 2000; Verdoit et al., 2003).

La première étude est consacrée à la pêcherie de la crevette royale par chalutage benthique et par pêche côtière. Pour la pêche côtière, les méthodes utilisées afin de déterminer les facteurs explicatifs de la variation de la production de ce crustacé dans le golfe de Gabès étaient les suivantes : une régression linéaire multiple sur les caractéristiques techniques des barques côtières (longueur, jauge brute et puissance motrice) pour identifier les variables les plus influentes, l'analyse des corrélations afin de détecter d'éventuels liens entre les variables, des tests paramétriques (ANOVA et ANCOVA) afin d'étudier l'effet de la composante spatiale et de l'effort de pêche sur la production de cette espèce et finalement une régression linéaire simple pour estimer la production à partir de la variable la plus explicative retenue. Le chalutage benthique dans le golfe de Gabès est une pêcherie plurispécifique et très hétérogène quant aux caractéristiques techniques des unités actives (Missaoui et al., 2000). Les stratégies de pêche peuvent différer d'une unité à une autre selon les caractéristiques techniques du chalutier, les espèces ciblées mais aussi selon la rentabilité économique attendue de la marée. La caractérisation de cette activité passe inévitablement par une typologie de cette

flotte. La typologie est la science de l'élaboration des types afin de faciliter l'analyse d'une réalité complexe (Biseau et Gondeaux, 1988). Une typologie a été implémentée sur la base de la production des différentes unités visités, de leurs caractéristiques techniques et des zones fréquentées. Cette typologie a été réalisée en 3 étapes: une Analyse en Composantes Principales (ACP), suivie d'une Classification Hiérarchique Ascendante (CHA), puis d'une segmentation. Une analyse canonique a été entreprise dans un second temps dans le but de dégager un modèle permettant d'affecter un chalutier à la zone d'activité. Une régression linéaire simple a été conduite afin de tester l'effet de la puissance motrice sur la production de la crevette royale.

Les grandes différences constatées pour l'activité du chalutage benthique, nous a amené à se focaliser davantage dans la seconde partie sur cette activité. Des enquêtes ont été effectuées dans le même port et le questionnaire rempli par les commandants de bord contenait des informations plus précises sur la zone d'activité (longitude et latitude), les caractéristiques techniques des chalutiers (longueur, puissance motrice, jauge brute et âge du chalutier, nombre de marins à bord, nombre de chaluts utilisés, nombre de jours de mer par marée) et les espèces capturées. Une analyse typologique a été réalisée sur la base de ces données afin d'identifier les différents métiers. Un métier définit l'activité du navire pendant la période de temps concernée, il rend compte des choix d'un engin, d'une espèce cible (ou d'un groupe d'espèces) voire d'un secteur (Biseau, 1994). Une approche cartographique a été également entreprise afin d'identifier les zones d'activités des différents métiers et la rentabilité économique de chaque métier a été évaluée.

L'identification de la répartition spatiale et l'évolution temporelle des activités de pêche sont des informations essentielles afin d'adapter les démarches d'aménagement sur la base des caractéristiques des régions. Une telle étude n'a jamais été élaborée et à cet effet nous avons essayé de caractériser les différentes activités à travers les espèces ciblées et les zones d'exploitation. La reconstitution de l'évolution des pêcheries est considérée comme un élément majeur pour mieux appréhender les dynamiques en cours ainsi que les enjeux de la gestion actuelle des pêches. Des méthodes de classification suivies de segmentation ont été utilisées pour suivre cette évolution et caractériser les différentes activités.

Les différents logiciels utilisés pour ces analyses étaient : Spad, SPSS, R et SAS pour les analyses statistiques et Mapinfo pour les cartographies.

3 Résultats et discussion

3.1 Pêche à la crevette royale *Penaeus kerathurus* (Forsk., 1775)

Les caractéristiques techniques des barques côtières (puissance motrice, jauge brute, longueur des unités) sont parfaitement corrélées et la puissance motrice semble être la variable la plus explicative de la production de la crevette royale. Cette dernière est influencée, entre autres,

par le nombre de pièces de trémail utilisé. L'effet de cet effort de pêche est amorti au delà de la valeur de 90 pièces, ainsi cette valeur correspondrait à l'effort de pêche optimum. La mise en évidence de cette valeur remarquable est d'une importance capitale pour une meilleure gestion de l'effort de pêche. En effet, tout accroissement de l'effort de pêche au delà de cette valeur ne se traduit que par une faible augmentation des débarquements. Le facteur abiotique considéré, la profondeur de la zone d'activité, n'influe pas sur la quantité des captures mais plutôt sur leurs qualités. Ainsi, conformément à la distribution de cette espèce mise en relief par Ben Khemis (1984), il s'en suit que les grosses crevettes occupent les profondeurs les plus importantes. A cet effet, les pêcheurs exploitant le large ont tendance d'amener les mêmes quantités de crevette mais avec une rentabilité économique plus intéressante. Les modèles finaux déterminant les variables explicatives de la production de la crevette royale ont montré l'importance de la puissance motrice de l'embarcation mais aussi le nombre de pièce de trémail utilisé. L'effet de ce dernier régresse à partir de la valeur de 90. Une limitation de l'effort de pêche en deçà de ce seuil est donc fortement recommandée en vue d'éviter le déploiement inefficace d'un excès de filets trémail.

Le chalutage benthique est une pêcherie plurispécifique. Des espèces indicatrices des zones ont été identifiées : la crevette mouchetée pour la zone interne du golfe de Gabès et les poissons pour la partie externe. La typologie établie a montré l'existence de deux stratégies de pêche différentes selon les espèces exploitées, la zone exploitée et les caractéristiques techniques des unités. Les unités les plus puissantes exploitent le large dont l'intérêt est de capturer des espèces à hautes valeurs marchandes destinées essentiellement à l'exportation contrairement aux unités les moins performantes qui exploitent la zone plus côtière du golfe. La production de la crevette royale par les chalutiers est indépendante de la stratégie de pêche choisie, ainsi il paraît que la crevette royale est une espèce vitale pour tous les chalutiers. Les débarquements de la crevette royale semblent dépendre étroitement de la puissance motrice de l'embarcation pour les unités exploitant la zone interne du golfe de Gabès.

3.2 Caractérisation de l'activité du chalutage benthique dans le golfe de Gabès

La typologie entreprise pour l'activité de chalutage benthique a fait ressortir quatre stratégies de pêche en fonction des caractéristiques techniques des unités et des espèces exploitées:

- Le premier métier formé par de vieilles unités aux caractéristiques techniques modestes qui exploitent la zone interne du golfe de Gabès et dont les espèces cibles sont les espèces côtières à savoir la crevette royale et la seiche.
- Le second métier englobe des chalutiers récents dont l'aire d'activité est plus étendue allant de 50 à 200 mètres et par suite ils n'ont pas d'espèces cibles en particulier.

- Le troisième métier est constitué par des unités de caractéristiques techniques performantes et d'un séjour en mer assez important dont les poissons représentent l'essentiel des captures.
- Le dernier métier est celui qui cible les espèces destinées à l'exportation à savoir la crevette rouge, le calmar, le poulpe et le petit poulpe. Il s'agit des unités congelât caractérisées par de grandes tailles et un nombre important de marins pour assurer les opérations de conservations des captures. Leurs aires d'activité est plutôt les fortes profondeurs, lieux de concentration de ces espèces.

L'approche économique montre que les dépenses dépendent des zones de pêches fréquentées. Le métier 4, formé des unités congelât, destiné à l'exportation est celui le plus efficace économiquement. A cet effet, on a pu remarquer un accroissement du nombre de ces unités au cours de la dernière décennie.

4 Évolution spatiale et temporelle des pêcheries du golfe de Gabès

Le suivi de l'évolution temporelle des pêcheries du golfe de Gabès fait ressortir les constatations suivantes :

- La production halieutique a connu trois phases. La pêche côtière était initialement l'activité la plus productive (1996, 1997, 1998), le chalutage benthique a pris de l'importance à partir de 1999 jusqu'à 2005 et finalement l'aquaculture a considérablement augmenté en 2006 et 2007.
- La flotte active, pour sa part, a connu trois étapes : la première est caractérisée par une augmentation considérable des thoniers (1996 à 1999), ensuite les senneurs et les chalutiers ont vu leurs effectifs s'accroître significativement et finalement une dernière phase (2005, 2006 et 2007) était caractérisée par la prédominance d'aucune flotte en particulier.
- L'essentiel de la main d'œuvre active du secteur était d'abord mobilisée pour la pêche côtière, puis la seconde phase était marquée par la dominance de la pêche aux poissons bleus pour finalement migrer vers la pêche au chalutage benthique et à la senne tournante.

Ainsi pour récapituler la dynamique de l'activité de pêche dans le golfe de Gabès, on pourra avancer les conclusions suivantes :

- Toute augmentation de la production d'un type de pêche a un effet significatif sur la mobilisation de la main d'œuvre. La main d'œuvre migre vers les activités les plus productives.
- La pêche côtière a subi l'effet de l'augmentation considérable de l'effort de pêche alloué au chalutage benthique ; ainsi la main d'œuvre et la production ont diminué.
- La production halieutique issue de la pêche est en train de céder sa place au bénéfice de l'aquaculture qui s'épanouit afin de faire face aux besoins croissants du marché national suite au déclin des stocks des poissons.
- La période stagnante des effectifs des unités montre une certaine méfiance des investisseurs à miser sur une activité suite à la diminution spectaculaire de la production halieutique.

Références

Allen R., Punsly R. (1984). Catch rates as indices of abundance of yellowfin tuna, *Thunnus albacares* in the Eastern Pacific Ocean. *Inter. Am., Trop. Tuna Comm. Bull.*, **18**, 303 - 379.

Ben Khemis L., 1984, Pêche à la crevette dans le golfe de Gabès. *Rapp. Doc. Inst. natn. scient. tech. Océanogr. Pêche. Salammbô*. 2, 3 - 30.

Ben Meriem S., Ezzedine-Najai S., Gharbi H., 1994a, Estimation des mortalités par pêche et analyse des rendements par recrue de la seiche *Sepia officinalis* (Céphalopode, Décapode) du golfe de Gabès. Conseil Général des pêches pour la Méditerranée. Consult. techn. Evaluation des stocks. Tunis, Tunisie, 8-12 nov. 1994. *Rapp. Pêches* FAO 533 suppl. : 231 - 245.

Ben Meriem S., Gharbi H., Ezzedine-Najai S., 1994b, Le rouget de roche (*Mullus surmuletus*) en Tunisie : évaluation des stocks et aménagement des pêcheries. Conseil Général des pêches pour la Méditerranée. Consult. techn. Evaluation des stocks. Tunis, Tunisie, 8-12 nov. 1994. *Rapp. Pêches* FAO 533 suppl., 269 - 284.

Ben Meriem S., Gharbi H., Bedoui R., 2005, Les pêcheries multispécifiques du golfe de Gabès. Etat d'exploitation des principales espèces benthiques et éléments d'aménagement. *7ème Forum Halieumétrique, Nantes* 21-22-23/6/2005. 7 pp.

Bertignac M., 1992, Les rendements par espèce de la pêche chalutière hauturière française de l'Atlantique Nord-Est entre 1972 et 1989. *Aquat. Living Ressour.* 5, 261 - 275.

- Bigelow K. A., Hampton, J., Miyabe, N., 2002, Application of a habitat-based model to estimate effective longline fishing effort and relative abundance of Pacific bigeye tuna (*Thunnus obesus*). *Fish. Oceanogr.* 11 (3), 143 - 155.
- Biseau A., Gondeau E., 1988, Apport des méthodes d'ordination en typologie des flottilles. *J. Cons. Int. Explor. Mer.* 44, 286 - 296.
- Biseau A., 1994, Usage des typologies. Communication au séminaire *IFREMER d'analyse de flottilles*. Nantes : 31 p.
- Ghorbel M., 1996, *Le Pageot commun Pagellus erythrinus (Poisson, Sparidae) écobiologie et état d'exploitation dans le golfe de Gabès*. Thèse de doctorat de Spécialité, 170 p. Fac. Sci. Sfax.
- Hattour A., 1991, Le chalutage dans les eaux tunisiennes, réalités et considérations législatives particulièrement dans les golfes de Tunis et de Gabès. *Notes. Inst. natn. scient. tech. Océanogr. Pêches, Salammbô* 1, 1 à 28.
- Hilborn R., Walters C., 1992, *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman & Hall, London.
- Jabeur C., Gobert B., Missaoui H., 2000, Typologie de la flottille de pêche côtière dans le golfe de Gabès (Tunisie). *Aquat. Living Resour.* 13, 421 - 428.
- Missaoui H., Jabeur Ch., Gobert B., Jarboui O., Elabed A., 2000, Analyse typologique de la flottille chalutière du golfe de Gabès (Sud Est Tunisie). *Bull. Inst. Scien. Tech. Mer Salammbô* 27, 15 à 26.
- Pelletier, D., Ferraris, J., 2000. A multivariate approach for defining fishing tactics from commercial catch and effort data. *Can. J. Fish. Aquat. Sci.*, 57 (1): 51 à 65.
- Verdoit, M., Pelletier, D. & Bellail, R., 2003. Are commercial logbook and scientific CPUE data useful for characterizing the spatial and seasonal distribution of exploited populations? The case of the Celtic Sea whiting. *Aquat. Living Resour.*, 16, 467 à 485.

Nonlinear FDA with Reproducing Kernels

Hachem KADRI *

* Adresse pour correspondance:
SequeL Team, INRIA Lille - Nord Europe
e-mail: hachem.kadri@inria.fr

Abstract

Recent statistical and machine learning studies have revealed the potential benefit of adopting a functional data analysis (FDA) point of view to improve learning when data are objects in infinite dimensional Hilbert spaces; however, nonlinear modeling of such data (aka functional data) is a topic that has not been sufficiently investigated, especially when response data are functions. Reproducing kernel methods provide powerful tools for nonlinear learning problems, but to date they have been used more to learn scalar or vector-valued functions than function-valued functions. Consequently, reproducing kernels for functional data and their associated function-valued RKHS have remained mostly unknown and poorly studied.

In this talk, I will describe a learning methodology for nonlinear FDA based on extending the widely used scalar-valued RKHS framework to the functional response setting. Then, I will introduce a set of rigorously defined reproducing operator-valued kernels suitable for functional response data, that can valuably applied to take into account relationships between samples and the functional nature of data (e.g. smoothness of the curves underlying the discrete observations). Finally, I will show experimentally that the nonlinear FDA framework is particularly relevant for speech and audio processing applications where attributes are really functions and dependent of each other.

References

Kadri, H., Duflos, E., Preux, P., Canu, S., and Davy, M. Nonlinear functional regression: a functional rkhs approach. In AISTATS 2010, JMLR: W&CP 9, pp. 111–125, Chia Laguna, Sardinia, Italy, 2010.

Kadri, H., Rabaoui, A., Preux, P., Duflos, E., and Rakotomamonjy, A. Functional regularized least squares classification with operator-valued kernels. In ICML 2011, Seattle, USA, 2011.

Analyzing functional data with lattice spatial dependence: A generalization of LISA map

Pedro DELICADO* and Sonia BRONER

* Adresse pour correspondance:
 Departament d'Estadística i Investigació Operativa
 Universitat Politècnica de Catalunya
 Despatx 218, Edifici C5, Campus Nord
 C/ Jordi Girona 1-3, 08034 Barcelona (SPAIN)
 e-mail: pedro.delicado@upc.edu

Abstract

The literature on spatial statistics for functional data is not extensive at this moment (see Delicado et al, 2010, for a review). Following Cressie (1993) for the definition of spatial processes and Ferraty and Vieu (2006) for that of functional random variables, we can define a *spatial functional process* as

$$\{\chi_s : s \in D \subseteq R^d\} \quad (4.1)$$

where s is a generic data location in the d -dimensional Euclidean space (d is usually equal to 2), the set $D \subseteq R^d$ can be fixed or random, and χ_s are *functional random variables*, defined as random elements taking values in an infinite dimensional space (or *functional space*). Typically χ_s , for every fixed s , is a real function from $[a, b] \subseteq R$ to R .

As it happens in univariate or multivariate spatial data analysis, the nature of the set D allows to classify spatial functional data. *Geostatistical functional data* appear when D is a fixed subset of R^d with positive volume and n points s_1, \dots, s_n in D are chosen to observe the random functions χ_{s_i} , $i = 1, \dots, n$. We say that we have a *functional marked point pattern*, when a complete function is observed at each point generated by a standard point process. *Functional areal data* (or *functional data in lattice*) correspond to the case of D being a fixed and countable set. Usually there is a bijection between D and a partition of a geographical area and, for any $s \in D$, χ_s is a summary function of an event happened at the part of the area corresponding to s by this bijection.

We focus on functional data with spatial dependence of lattice type. An alternative less formal way of introducing them is as follows. The geographical unit under study (a region,

a country, the world, ...) is divided in smaller sub-areas (mainly because administrative reasons) and a functional data is observed at each sub-area. Spatial dependence arises when the functional data observed in a given sub-area are more similar to those observed in the neighboring areas than those corresponding to remote areas. The aims for this kind of spatial functional data analysis are similar to those concerning univariate or multivariate areal data: detection of the spatial dependence (via spatial autocorrelation testing), identification of spatial clusters, and to modeling the spatial dependence (via spatial regression models, for instance) are probably the main ones. Thus it is natural to adapt descriptive and inferential techniques used for univariate and multivariate areal data to the case of having a function at each area. LISA map (Anselin, 1995) is one of the descriptive tools most extensively used in spatial univariate data analysis of lattice type.

In Delicado and Broner (2008) we proposed a first LISA map generalization based on the distances between the features observed in different areas. Therefore it can be applied when the observations are functional data. In the present work we present a refined and more powerful version of the distance based LISA map. As an illustration we analyze the population pyramids (considering them as density functions) of all countries in the world. Our first aim is to detect whether there is spatial dependence between population pyramids or not. In addition to that, a modification of the proposed algorithm allows us to detect spatial clusters of neighboring countries with extremely similar population pyramids.

To define the distance between population pyramids we take into account that density functions are compositional data with infinite dimension (see Delicado, 2011). So the distance between two densities f and g is defined as

$$d(f, g) = \left[\frac{1}{2(b-a)} \int_I \int_I \left(\log \frac{f(x)}{f(y)} - \log \frac{g(x)}{g(y)} \right)^2 dx dy \right]^{1/2},$$

where $I = [a, b]$ is the interval where f and g are defined.

Références

- Anselin, L. (1995). Local indicators of spatial association-lisa. *Geographical Analysis* **2**, 93-115.
- Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.
- Delicado, P. (2011) Dimensionality reduction when data are density functions. *Computational Statistics and Data Analysis*, **55**, 401-420. doi: 10.1016/j.csda.2010.05.008
- Delicado, P. and S. Broner (2008). Distance-based LISA maps for multivariate lattice data.

Technical Report <http://hdl.handle.net/2117/2421>, Universitat Politècnica de Catalunya.
Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010) Statistics for spatial functional data. *Environmetrics*. **21**, 224-239. doi: 10.1002/env.1003
Ferraty, F. and P. Vieu (2006). *Non Parametric Functional Data Analysis*. Theory and Practice. Springer.

Support Vector Machine Quantile Regression approach for Functional Data

Yousri HENCHIRI *

(work in collaboration with Christophe CRAMBES and Ali GANNOUN)

* Adresse pour correspondance:

Institut de Mathématique et de Modélisation de Montpellier

Equipe de Probabilités et Statistique

Université Montpellier II

Place Eugène Bataillon cc051

34095 Montpellier cedex, France.

e-mail: henchiri@math.univ-montp2.fr

Résumé

Investigating the relationship between a response Y and a covariate X is a core issue in statistics. There has been a vast literature dealing with regression on conditional mean function by exploiting all kinds of methods such as parametric, nonparametric and semi-parametric approaches. However, in data processing, such information cannot be obtained by classical statistical models that deal only with the conditional mean function estimation. The disadvantage of this classical regression is that it is sensitive to outliers and may be inappropriate in some cases, as when the distribution is multimodal or highly asymmetric. These problems can be solved by using another useful descriptive statistic which is robust to heavy-tailed error distributions and outliers, called *conditional quantiles*. Conditional quantiles can provide a more complete picture of the conditional distribution of Y given $X = x$ at different quantile levels. For the above reasons, conditional quantiles are used in many areas of applied research and are frequently used in a regression setup (called *quantile regression*) [Koenker (2005)]. Usually, conditional quantiles estimation can be derived by two ways :

- either inverting a conditional distribution function,
- or solving an optimization problem.

In this context, two main alternatives can be considered. The first alternative is to use a parametric approach via the linear model for conditional quantiles, first introduced by Koenker and Bassett (1978). However, and in many practical situations, this linear quantile approach may not be rich enough to capture the underlying relationship between the quantile of the response Y and its associated covariate X . The other alternative to estimate conditional quantiles is the nonparametric approach or nonparametric smoothing techniques [Chaudhuri (1991) and De Gooijer et al. (2002)] which constitute a branch of modern statistics intending to fit a much larger class of models to reduce modeling biases. In our work, we will focus on the last approach.

In the last decades, we have witnessed an explosion in the mass of available data through the development of media, where many computers can provide large amount of complex information. Such data, for example collected and saved in less and less time, are now well known as *Functional Data* which are random variables observed in some continuous way. In this context, several researchers have focused on quantifying a new relationship between the response Y and the functional covariate $X(t)$. The failure of the standard multivariate statistical analysis to handle functional data has spurred into the development of new methodologies, able to extract relevant information from functional data sets, grouped under the name of *Functional Data Analysis (FDA)*.

In this talk, we explain, in the field of **FDA**, the statistical learning theory and the classical regularization theory for function approximation through *Support Vector Machine (SVM)* method [Vapnik (1998), Schölkopf and Smola (2002) and Steinwart and Christmann (2008)]. The learning criterion is based on a regularized risk for which the complexity of the model is measured by a norm in *Reproducing Kernel Hilbert Space (RKHS)* [Berlinet and Thomas-Agnan (2004) and Steinwart and Christmann (2008)]. First the functional observations, $X_1, \dots, X_n \in \mathcal{X} := \mathbf{L}^2([0, 1])$, are projected onto a finite dimensional subspace having a suitable orthonormal system so that the projections of the X_i 's can be identified with elements of $\mathbb{R}^d = \tilde{\mathcal{X}}$. Next ordinary *Support Vector Machine Quantile Regression* approach using *Iterative Reweighted Least Squares* procedure to solve the penalized minimisation problem which has no explicit solution is applied to the projected data, which is possible because the first step reduces the functional data setting to a non-functional one. Then we establish weak consistency of this estimator. The various parameters needed for the construction of this estimator are automatically selected by data-splitting and by penalized empirical risk minimization. The performances of the proposed algorithm, in terms of forecast accuracy, are evaluated with simulation and real data set studies.

Références

- Aronszajn, N. (1950) Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68**(3), 337–404.
- Berlinet, A. and Thomas-Agnan, C. (2004) *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publisher.
- Chaudhuri, P. (1991) Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, **39**(2), 246–269.
- Crambes, C., Gannoun, A. and Henchiri, Y. (2011) Weak consistency of the Support Vector Machine Quantile Regression approach when covariates are functions. *Statistics and Probability Letters*, **81**(12), 1847–1858.
- Crambes, C., Gannoun, A. and Henchiri, Y. (2012) Support Vector Machine Quantile Regression approach for Functional Data : simulation and application studies. *Soumis*.
- De Gooijer, J.G., Gannoun, A. and Zerom, D. (2002) Mean squared error properties of kernel-based multi-stage conditional median predictor for time series. *Statistics and Probability Letters*, **56**, 51–56.
- Ferraty, F., Rabhi, A. and Vieu, P. (2005) Conditional quantiles for dependent functional data with application to the climatic EL Ni no phenomenon. *Sankhya: Special Issue on Quantile Regression and Related Methods*, **67**, 378–398.
- Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker, R. (2005) *Quantile regression*. Cambridge University Press, Cambridge.
- Li, Y., Liu, Y. and Zhu, J. (2007) Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255–268.
- Preda, C. (2007) Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference*, **137**(3), 829–840.
- Rossi, F. and Villa-Vialaneix, N. (2006) Support Vector Machine for Functional Data Classification. *Neurocomputing*, **69**(7-9), 730–742.
- Schölkopf, B., Smola, A. J. (2002) *Learning with Kernels*. MIT Press.
- Steinwart, I. and Christmann, A. (2008) *Support Vector Machines*. Springer, New York.
- Vapnik (1998) *Statistical Learning Theory*. John Wiley and Sons.

Caractérisation d'événements à partir de signaux expérimentaux

Laurent PANTERA*

* Adresse pour correspondance:
Département d'Etudes des Réacteurs
Service de Réalisations d'Essais en Sûreté
Laboratoire de Préparation et Réalisation des Essais
Commissariat à l'Energie Atomique
CEA Cadarache, 13108 Saint-Paul-Lez-Durance
laurent.pantera@cea.fr

Mots clefs : nucléaire, combustible, traitement du signal, classification, diagnostic

Résumé

CABRI est un réacteur expérimental destiné aux études de sûreté en soutien au parc électronucléaire. Le travail présenté entre dans le cadre du programme d'essais "CABRI International Program (CIP)". Celui-ci a pour objectif d'étudier le comportement des éléments de combustible d'un Réacteur à Eau sous Pression ou REP (PWR Pressurized Water Reactor en anglais) à haut taux de combustion lorsqu'ils sont soumis à un accident d'insertion de réactivité correspondant à l'éjection d'une barre de contrôle. Chaque essai porte sur l'étude de comportement d'un seul élément combustible appelé également crayon. Pour fixer les idées, celui-ci est constitué d'un ensemble de pastilles d'oxyde d'uranium empilées sur une hauteur d'un mètre environ et placé dans une gaine en alliage de zirconium de 9.5 mm de diamètre et de 0.57 mm d'épaisseur. Les essais consistent à soumettre l'élément combustible à des puissances pouvant atteindre les 20 GW. Ces montées en puissance sont réalisées sur de très courtes durées (10 à 100 ms). Après chaque essai, les expérimentateurs sont amenés dans le cadre d'un rapport préliminaire à se prononcer sur l'absence ou pas de rupture de la gaine de l'élément combustible testé. Ce diagnostic est effectué à partir des signaux obtenus en ligne, notamment à partir de deux microphones (fréquence d'acquisition égale à 1 MHz) placés en amont et en aval du dispositif d'essai. Cet exposé souhaite montrer comment au sein d'une équipe d'expérimentateurs très mobilisés autour de l'élaboration d'expériences, il a été possible grâce à la souplesse du langage R de réorganiser les mesures de nos anciens

essais pour pouvoir les analyser et les utiliser dans nos futures prévisions. Nous avons dans un premier temps récupéré tous les signaux suivis en ligne pour chaque portion d'une seconde d'essai [1] et effectué une découpe de ces secondes expérimentales en événements en se basant sur la variation de la variance [2] des signaux des microphones. Chaque événement a été stocké dans une base de données relationnelle (PostgreSQL) [3] puis nous avons extrait de ces signaux des indicateurs temporels et fréquentiels sur lesquels nous avons pu effectuer des analyses en composantes principales et des classifications [4]. L'agencement de ces méthodes a pu ensuite être intégré dans une interface développée en JAVA [5], langage utilisé en interne pour effectuer le suivi en temps réel des essais. Nous cherchons maintenant à affiner nos classifications pour mettre en évidence des éléments précurseurs aux phénomènes de ruptures.

Références

- [1] Zeileis, A., Grothendieck, G., Ryan, J., Andrews, F. Package zoo, <http://zoo.R-Forge.R-project.org/>, consulté le 24 avril 2012
- [2] Scrucca, L.(2004). qcc: An R package for quality control charting and statistical process control, R news 4:11-17
- [3] Urbanek, S. Package RJDBC, <http://www.rforge.net/src/contrib/Documentation/RJDBC.pdf>, consulté le 24 avril 2012
- [4] Chessel, D., Dufour, A.-B. and Thioulouse J. (2004). The ade4 package - I : One-table methods, R news 4:5-10
- [5] Urbanek, S. Package rjava, <http://www.rforge.net/rJava/>, consulté le 24 avril 2012

Testing for a change in covariance operator

Daniela JARUŠKOVÁ*

* Adresse pour correspondance:
Czech Technical University
Faculty of Civil Engineering, Thákurova 7
Prague, Czech Republic
e-mail: jarus@mat.fsv.cvut.cz

Résumé

A test procedure for testing equality of two covariance operators against the alternative that the operators are different in a principal components approximation is suggested. Asymptotic distribution of the proposed test statistic is presented. It is shown how the bias of the test statistic depends on the eigenvalues spacings in a case of finite sample size.

References

- [1] Antoch J. and Hušková M.: Permutation tests in change point analysis. *Statistics and Probability Letters* 53 (2001), 37–46.
- [2] Aue A., Gabrys R., Horváth L., and Kokoszka P.: Estimation of change-point in the mean function of functional data. *Journal of Multivariate Analysis* 100 (2009), 2254–2269.
- [3] Benko M., Härdle W., Kneip A.: Common functional principal components. *Annals of Statistics* 37 (2009), 1–34.
- [4] Bosq D.: *Linear processes in functional spaces*. Springer, New York 2000.
- [5] Dauxois J., Pousse A. and Romain Y.: Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis* 12 (1982), 136–154.
- [6] Ferraty F. (editor): *Recent advances in functional data analysis and related topics*. Physica-Verlag, Berlin-Heidelberg 2011.
- [7] Hall P. and Hosseini-Nasab M.: On properties of functional principal components analysis. *J. R. Stat. Soc. Ser B* 68 (2006), 109–126.

- [8] Hall P. and Hosseini-Nasab M.: Theory for high-order bounds in functional principal components analysis. *Math. Proc. Camb. Phil. Soc.* 146 (2009), 225–256.
- [9] Horváth L. and Kokoszka P.: *Inference for functional data with applications*. Springer, Heidelberg 2012.
- [10] Panaretos V.M., Kraus D., Maddocks J.H.: Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *JASA* 105 (2010), 670–682.