

## Editorial of the Special Issue “Functional data”

**Title:** Editorial du numéro spécial “Données fonctionnelles”

Sophie Lambert-Lacroix<sup>1</sup> et André Mas<sup>2</sup>

Functional data analysis aims to study and model observations which are by nature not vectors but random curves. For instance, the observation over  $n$  days of the price of share  $A$  provides a dependent sample of size  $n$  of the random function “Daily rate for share  $A$ ”. The same is true for the recording of temperatures at a given location or the electricity consumption of a country. Each of these curves is of course initially discretized and can therefore be represented by a vector. But the limitations of the classic multivariate approach in this context quickly become apparent if the discretization frequency is too low or too high, or if there are corrupted data. It is in this context that functional data modeling has been a very active research area over the last twenty years. There has been remarkable development in this field, leading to modern advances in information technology and the advent of “Big Data”. Specific tools from signal processing or functional analysis must often be implemented and added to the statistician’s large collection of tools to circumvent the theoretical and numerical problems which then appear.

This special issue attempts to give the broadest possible overview of the fields of application of statistics for functional data. Five of the six articles were written at least in part by young researchers. The reader will be able to note the diversity of themes and models (dependent or independent data, nonparametric methods, or Bayesian spatial or time series), but also application domains : Oceanography, chemistry, agronomy, economics...

Frédéric Ferraty offers a thorough overview of the main regression models for functional data with a scalar response and a functional predictor. He gives some recommendations on how to put them to good use. This methodological approach is illustrated on spectrometric data. Séverine Bayle, Pascal Monestiez and David Nérini also address regression in their article, in which they consider a model where the variable to be predicted and the predictor are both curves. They propose a methodological approach through an application of this model on environmental data. Camille Ternynck used a kernel estimator for nonparametric regression with a variable response presenting spatial dependence and a functional predictor. This estimator is proved to be mean square consistent and its good behavior is illustrated from numerical experiments on simulated

<sup>1</sup> UJF-Grenoble 1 / CNRS / UPMF / TIMC-IMAG UMR 5525, Grenoble, F-38041, France.

E-mail : [sophie.lambert@imag.fr](mailto:sophie.lambert@imag.fr)

<sup>2</sup> UMR CNRS 5149, Institut de Modélisation Mathématiques de Montpellier, Case courrier 051 34095 Montpellier cedex 5 - France.

E-mail : [andre.mas@univ-montp2.fr](mailto:andre.mas@univ-montp2.fr)

data. Matthieu Saumard proposes a new model for estimating equations in an article with a strong theoretical identity. He proposes an estimator for the interest parameter of this model and studies its asymptotic properties. Damien Juéry, Christophe Abraham and Bénédicte Fontez consider the cluster analysis of curves. They successfully combined functional data and Bayesian methods to obtain a promising clustering method. Finally Jairo Cugliari, Xavier Brossat, Anestis Antoniadis and Jean-Michel Poggi tackle the complex problem of the forecast of electricity consumption by highlighting new breakthroughs. They consider a prediction model for functional nonstationary time series and apply it to the electricity demand of Electricité de France (EDF).

---

La statistique pour données fonctionnelles s'attache à étudier et modéliser des observations qui, par nature, ne sont pas des vecteurs mais des courbes aléatoires. Par exemple l'observation pendant  $n$  journées du cours d'une action  $A$  fournit un échantillon dépendant de taille  $n$  de la fonction aléatoire "cours journalier de l'action  $A$ ". Il en est de même de l'enregistrement des températures en un lieu donné ou de la consommation d'électricité d'un pays. Chacune de ces courbes est bien entendu initialement discrétisée donc représentable par un vecteur. Mais l'approche multivariée classique dans ce contexte révèle rapidement ses limites si la fréquence de discrétisation est trop faible, ou trop élevée, ou si des erreurs de mesures entachent les observations. C'est dans ce cadre que la modélisation des données fonctionnelles a connu au cours des vingt dernières années un développement remarquable, accompagnant -à l'instar de bien d'autres domaines des statistiques modernes- les progrès de l'informatique et l'avènement du *Big Data*. Des outils spécifiques, issus du traitement du signal ou de l'analyse fonctionnelle, doivent souvent être mis en œuvre et ajoutés dans l'arsenal du statisticien pour contourner les problèmes théoriques et numériques qui apparaissent alors.

Ce numéro spécial tente de donner un aperçu aussi large que possible des champs d'application de la statistique pour données fonctionnelles. Sur les six articles, cinq ont été rédigés au moins en partie par de jeunes chercheurs. Le lecteur constatera la diversité des thématiques et des modèles (données dépendantes ou indépendantes, méthodes non-paramétriques ou bayésiennes, processus spatiaux ou temporels), mais aussi des domaines d'application : océanographie, agronomie, chimie, économie...

Ainsi Frédéric Ferraty propose un survol commenté des principaux modèles de régression pour données fonctionnelles pour lesquels la réponse est réelle et le prédicteur fonctionnel et il donne quelques recommandations pour utiliser ces modèles voire les combiner à bon escient. Cette démarche est illustrée sur des données de spectrométrie. Il est également question de régression dans l'article de Séverine Bayle, Pascal Monestiez et David Nérini qui considèrent un modèle où la variable à prédire et la variables prédictive sont des courbes. Ils présentent une démarche méthodologique à travers un exemple d'application de ce modèle sur des données environnementales. Camille Ternynck étudie via un estimateur à noyau des données de nature originales où la variable réponse est spatiale et la variable prédictive est une courbe. Elle montre les propriétés de convergence de cet estimateur et illustre son bon comportement via des simulations.

Matthieu Saumard introduit un nouveau modèle d'équation estimantes au travers d'un article aux contours théoriques marqués. Il propose un estimateur du paramètre d'intérêt et étudie ses propriétés asymptotiques. Damien Juéry, Christophe Abraham et Bénédicte Fontez considèrent le problème de classification de courbes non supervisée. Ils mêlent avec succès les données fonctionnelles et les méthodes bayésiennes pour obtenir une méthode de classification non supervisée prometteuse. Enfin Jairo Cugliari, Xavier Brossat, Anestis Antoniadis et Jean-Michel Poggi se penchent sur la problématique complexe de la prévision de la consommation d'électricité en mettant en lumière de nouvelles avancées. Ils considèrent un modèle de prévision pour des séries chronologiques fonctionnelles non stationnaires et des applications à la demande d'Electricité de France (EDF).