

# Offre de postdoctorat en statistique

## Présentation

**Sujet.** Noyaux reproduisants caractéristiques pour des données fonctionnelles

**Durée.** 15 mois.

**Lieu.** Université Rennes 2/ Laboratoire IRMAR

## Description du postdoc

Les espaces de Hilbert à noyau reproduisant (RKHS) constituent un cadre théorique puissant et flexible qui a profondément transformé l'analyse statistique et l'apprentissage automatique [1, 2, 3]. Leur principal atout réside dans leur capacité à capturer des relations complexes et non linéaires entre les données tout en conservant une formulation linéaire dans un espace de plus grande dimension [2]. Grâce aux noyaux caractéristiques, les RKHS permettent de représenter des distributions de probabilité de manière unique, facilitant ainsi des tâches fondamentales comme les tests à deux échantillons avec la distance du Maximum Mean Discrepancy (MMD) [4], ou les tests d'indépendance avec le critère HSIC [5]. En apprentissage, les RKHS jouent un rôle clé pour la classification SVM [6], la régression [7] et la réduction de dimension [8]. Leur utilisation permet de s'affranchir de l'hypothèse de linéarité souvent trop restrictive et d'exploiter efficacement des structures complexes de données comme les graphes et textes.

Dans ce postdoc, nous nous intéressons à la construction de noyaux caractéristiques pour des données fonctionnelles. En effet, une donnée fonctionnelle est une réalisation d'une variable aléatoire fonctionnelle représentant des courbes, des images ou d'autres formes de données à structure continue. Mathématiquement, une donnée fonctionnelle est une réalisation d'une variable aléatoire à valeurs dans un espace de dimension infinie<sup>1</sup>. L'utilisation de ce type de données devient de plus en plus courant grâce aux avancées technologiques en matière de stockage et de calcul. Des disciplines variées, comme la médecine [10], la biologie [11], la climatologie [12] ou encore l'économétrie [13], génèrent fréquemment ce type de données. Le traitement des données fonctionnelles requiert des méthodes adaptées, distinctes des outils classiques de la statistique multivariée et nécessitant des ajustements spécifiques pour tenir compte de leur complexité.

Bien que les espaces de Hilbert à noyau reproduisant (RKHS) et les noyaux caractéristiques aient été largement étudiés en statistique et en apprentissage automatique, leur application aux variables aléatoires fonctionnelles reste encore un champ peu exploré. En effet, la grande majorité des travaux existants se concentrent sur des variables aléatoires en dimension finie, tandis que l'extension aux données de nature fonctionnelle pose plusieurs défis théoriques. La principale difficulté réside dans la complexité des espaces fonctionnels, qui impose l'utilisation de noyaux respectant des contraintes de régularité et de stabilité. Malgré ces défis, quelques

---

1. Suivant le contexte, une structure peut être imposée à cet espace, comme celle d'un espace de Hilbert réel séparable.

études ont exploré cette direction, bien que leur nombre demeure très limité. L'article notable [14], propose une généralisation du noyau gaussien au cadre fonctionnel et établit son caractère caractéristique. Un autre travail à mentionner est [15], où les auteurs construisent un noyau caractéristique (et universel) sur un espace de trajectoires en s'appuyant sur la notion de signature robuste.

À la lumière de ces travaux, deux axes complémentaires sont à explorer. Le premier consiste à étendre les noyaux caractéristiques existants en dimension finie aux espaces fonctionnels. Le second vise à s'appuyer sur les travaux récents pour concevoir de nouvelles familles de noyaux adaptés aux données fonctionnelles. Enfin, la question de l'impact de la représentation discrète des données fonctionnelles sur les performances de la MMD constitue un aspect à approfondir.

La bibliographie ci-dessous donne quelques éléments indicatifs.

## Candidature et contact

**Profil du candidat.** Un doctorat en mathématiques appliquées est requis, avec de solides connaissances en statistique.

**Candidatures.** Les candidatures doivent être soumises par voie électronique aux adresses ci-dessous. Elles doivent contenir les documents suivants :

- un curriculum vitæ,
- une lettre de motivation.

### Contacts.

**Magalie Fromont**  
IRMAR (UMR CNRS 6625)  
Université Rennes 2  
[magalie.fromont@univ-rennes2.fr](mailto:magalie.fromont@univ-rennes2.fr)

**Madison Giacofci**  
IRMAR (UMR CNRS 6625)  
Université Rennes 2  
[joyce.giacofci@univ-rennes2.fr](mailto:joyce.giacofci@univ-rennes2.fr)

**Anouar Meynaoui**  
IRMAR (UMR CNRS 6625)  
Université Rennes 2  
[anouar.meynaoui@univ-rennes2.fr](mailto:anouar.meynaoui@univ-rennes2.fr)

## Références

- [1] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3) :337–404, 1950.
- [2] Bernhard Schölkopf and Alexander J Smola. *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.
- [3] Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, 2011.
- [4] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13 :723–773, 2012.
- [5] Arthur Gretton, Olivier Bousquet, Alexander Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International Conference on Algorithmic Learning Theory*, pages 63–77. Springer, 2005.

- [6] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, 1995.
- [7] Vladimir Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [8] John Shawe-Taylor and Nello Cristianini. Kernel principal component analysis and its applications in face recognition and active learning. *Advances in Kernel Methods*, pages 315–328, 2004.
- [9] Francis Bach, Gert R G Lanckriet, and Michael I Jordan. Learning with submodular functions : A convex optimization perspective. In *Advances in Neural Information Processing Systems*, pages 43–50, 2004.
- [10] Fabrizio Maturo and Rosanna Verde. Pooling random forest and functional data analysis for biomedical signals supervised classification : Theory and application to electrocardiogram data. *Statistics in Medicine*, 41(12) :2247–2275, 2022.
- [11] Guangzhe Fan, Jiguo Cao, and Jiheng Wang. Functional data classification for temporal gene expression data with kernel-induced random forests. In *2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 1–5. IEEE, 2010.
- [12] Daniela Ballari, Ramón Giraldo, Lenin Campozano, and Esteban Samaniego. Spatial functional data analysis for regionalizing precipitation seasonality and intensity in a sparsely monitored region : Unveiling the spatio-temporal dependencies of precipitation in ecuador. *International Journal of Climatology*, 38(8) :3337–3354, 2018.
- [13] Rituparna Sen and Claudia Klüppelberg. Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry*, 35(4) :1028–1043, 2019.
- [14] George Wynne and Andrew B Duncan. A kernel two-sample test for functional data. *Journal of Machine Learning Research*, 23(73) :1–51, 2022.
- [15] Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, 23(176) :1–42, 2022.