



## *Master internship position*

---

*Topic: Modeling space-time series with kernels*

### ***Advisor***

Florence d'Alché-Buc

### ***Location***

Telecom ParisTech, 46 rue Barrault, 75013 Paris

### ***Starting time***

April/May 2017

### ***Research group***

IDS dept, S2A group

### ***Key-words***

Spatio-temporal data, functional time series, operator-valued kernels, Reproducing Kernel Hilbert Spaces, Random Fourier Feature Operators

### ***Internship Subject:***

Spatio-temporal data are ubiquitous in many application areas such as climate science, transportation, epidemiology, internet of things, and social networks to name but a few. These data accumulate over time and rapidly grow into a corpus of massive datasets. Dealing with both spatial and temporal aspects raises exciting challenges in Machine Learning. This internship is concerned with setting up a novel framework for modeling spatio-temporal data through the modelling of functional time series. The main research direction is to take benefit from working in Reproducing Kernel Hilbert Spaces of function-valued functions, using operator-valued kernels. The intern will devote his/her time to study how to extend all the fundamental notions of time-series modelling in the context of spatio-temporal data and time series of functions of space. For that purpose, he/she will rely on the growing literature on functional statistics (see Ramsay's book) and the body of work concerning operator-valued kernels (Seneke and Tempel'man 1972, Michelli and Pontil 2005, Alvarez et al. 2012), input output regression (Brouard et al. 2016) and function-valued regression (Kradi et al. 2015).

Besides a general framework based on operator-valued kernels, the intern will also be aware of the crucial necessity of numerical efficiency and will rely on approximation of kernels and feature maps to build efficient algorithms. Applications on public datasets and industrial datasets from partners of the ML4bigData Chair will be encouraged.

## *La Chaire Machine Learning for Big Data*

Le traitement statistique des masses de données convoque à la fois mathématiques appliquées et informatique, à travers une discipline en plein essor : le Machine Learning ou apprentissage statistique.

Créée en septembre 2013 avec le soutien de la Fondation Télécom et financée à hauteur de près de 2 M€ par quatre entreprises partenaires : Criteo, PSA Peugeot Citroën, Safran et BNP Paribas, la Chaire Machine Learning for Big Data est portée par le mathématicien Stéphan Cléménçon, Enseignant-Chercheur, Professeur au sein du Département du Traitement du Signal et des Images à Télécom ParisTech.



**BNP PARIBAS**  
La banque d'un monde qui change

**criteo**

**PSA**  
GROUPE

**SAFRAN**  
AEROSPACE · DEFENCE · SECURITY

**Valeo**  
SMART TECHNOLOGY  
FOR SMARTER CARS

Proposant cinq axes de recherche méthodologiques, enrichis par des applications industrielles concrètes, cette Chaire a pour objectif d'animer, en interaction avec ses partenaires, une activité de recherche de pointe en Machine Learning, ainsi que de proposer des programmes de formation.

La variété des données aujourd'hui disponibles (nombres, images, textes, signaux), leur grande dimension et leur volumétrie rendent souvent inopérantes les méthodes statistiques traditionnelles reposant sur le prétraitement humain et un long travail de modélisation. Le Machine Learning vise donc à élaborer et étudier des algorithmes, à vocation prédictive le plus souvent, permettant à des machines d'apprendre automatiquement à partir des données et à effectuer des tâches de façon performante.

Les avancées technologiques, l'omniprésence des capteurs (systèmes embarqués, objets connectés, Internet...) et l'explosion des réseaux sociaux s'accompagnent d'un véritable déluge de données, propulsant les sciences de l'information au centre du processus de valorisation des masses de données. En plus de la collecte et du stockage, l'enjeu est de pouvoir analyser ces données afin d'optimiser les décisions et mettre au point de nouvelles applications.

Au-delà du buzz médiatique dont il fait l'objet, le Big Data est donc un sujet stratégique majeur, au cœur d'enjeux économiques et sociétaux considérables. Son impact est désormais perçu dans presque tous les secteurs de l'activité humaine : de la recherche scientifique à la médecine en passant, entre autres, par la finance, le bâtiment, l'e-commerce, la défense ou les transports.

En savoir plus sur la Chaire, ses axes de recherche, ses activités, ses publications :

<http://machinelearningforbigdata.telecom-paristech.fr>

### *Candidate profile*

Master Student (year 2)

- Machine Learning / Statistics
- Strong background in mathematics (functional analysis, Hilbert spaces)
- Good skills in optimisation and in programming

### *Candidatures*

To be sent to [florence.dalche@telecom-paristech.fr](mailto:florence.dalche@telecom-paristech.fr)

- Curriculum Vitae, Motivation letter (directly in the email)
- Grades report of the two previous years + Name of a reference professor

## *Bibliography and references*

Mauricio A. Álvarez, Lorenzo Rosasco, and Neil D. Lawrence. Kernels for Vector-Valued Functions: A Review. *Found. Trends Mach. Learn.* 4, 3 (March 2012), 195-266.

Alexander Aue and Diogo Durbant Norinho and Siegfried Hörmann}, On the Prediction of Stationary Functional Time Series, *Journal of the American Statistical Association*, 110:509, 2015.

Romain Brault, Markus Heinonen, Florence d'Alché-Buc, Random Fourier Features for Operator-valued kernels, *ACML 2016*.

Romain Brault, Néhémym Lim, Florence d'Alché-Buc, Scaling up Vector Autoregressive models with operator-valued Random Fourier features, *AALTD'16, joint workshop to ECML/PKDD 2016*, to appear in *LNCS*.

Céline Brouard, Marie Szafranski, Florence d'Alché-Buc, Input Output Kernel Regression: Supervised and Semi-Supervised Structured Output Prediction with Operator-Valued Kernels *17(176):1-48*, (2016).

Carmeli, Claudio, et al. 'Vector valued reproducing kernel Hilbert spaces and universality.' *Analysis and Applications* 8.01 (2010): 19-61.

Ciliberto, Carlo, et al. 'Convex learning of multiple tasks and their structure.' *International Conference on Machine Learning (ICML)*. 2015.

Dinuzzo, Francesco. 'Learning output kernels for multi-task problems.' *Neurocomputing* 118 (2013): 119-126.

Ferraty, Frédéric, and Philippe Vieu. *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media, (2006).

Néhémym Lim, Florence d'Alché-Buc, Cédric Auliac, George Michailidis: Operator-valued kernel-based-vector autoregressive models for network inference. *Machine Learning* 99(3): 489-513 (2015)

Rob J. Hyndman, Han Lin Shang, Forecasting functional time series, *Journal of the Korean Statistical Society*, Volume 38, Issue 3, September 2009, Pages 199-21, (2009).

Ruiz-Medina, M.D., Espejo, R.M., Ugarte, M.D. et al., Functional time series analysis of spatio-temporal epidemiological data, *Stoch Environ Res Risk Assess* (2014) 28: 943.

Maxime Sangnier, Olivier Fercoq, Florence d'Alché-Buc, Joint Quantile Regression in vector-valued RKHSs, *NIPS 2016*.

E. Senkne and A. Tempel'man. Hilbert spaces of operator-valued functions. *Lithuanian Mathematical Journal*, 13(4):665-670, 1973.

Ramsay, J. and Silverman, B. W. *Functional Data Analysis*, Springer series in statistics, 2005.