Steven **Golovkine**

PhD in Applied Mathematics, Statistics

(+33) 06 72 28 22 71 | steven_golovkine@icloud.com | stevengolovkine.netlify.app | StevenGolovkine | Steven Golovkine

Current Position

Postdoctoral Researcher on the FAST project

Limerick, Ireland

University of Limerick — *Feb. 2022 - Present*

- Develop novel, computationally efficient statistical models and algorithms for the modelling of multivariate sensor data.
- Demonstrate increased computational efficiency, so that the methodologies are applicable to modern, large-scale datasets.
- Supervisor: Norma Bargary (UL) and Andrew Simpkin (NUI Galway)

Education

PhD. in Applied Mathematics, Statistics

Guyancourt, France

Technocentre, Renault and CREST (Ensai) — *Jan. 2018 - Jun. 2021*

- **Title:** Statistical methods for multivariate functional data
- **Supervisors:** Valentin Patilea (Ensai, CREST), Nicolas Klutchnikoff (Univ Rennes, IRMAR)
- **Funding:** Partnership with Groupe Renault through a CIFRE convention.
- **Abstract:** The topic of this thesis is related to functional data analysis and is motivated by modern data from automobile industry. The standard functional data methods rely on the assumption that the curves are continuously observed, without error. However, in general, the real data is neither continuously nor exactly observed. Therefore, a crucial step is to recover the trajectories from noisy measurements at discrete random points. For that, we propose an original point of view: the local regularity of the process generating the curves. Thus, combining information both within and across trajectories, we propose a simple estimator for this local regularity. Given this estimate, we build a nearly optimal local polynomial smoother of the curves from a sample of noisy trajectories. Nonparametric estimators for the mean and the covariance functions of functional data, using the local regularity of the process, are derived. Moreover, we propose a model-based clustering algorithm for a general class of functional data for which the components could be curves or images. Results of both simulated and real data show the good performances of this method. A Python package, implementing the methods and publicly available, has been developed.
- **Keywords:** Adaptive optimal smoothing, functional data analysis, gaussian mixtures, Hölder exponent, local polynomials, model-based clustering, multivariate functional principal component analysis, traffic flow.
- **Defense**: June 18, 2021 in front of the jury composed of André Mas (Professor, Montpellier University, President), Sophie Dabo-Niang (Professor, Lille University, Reviewer), Alois Kneip (Professor, Bonn University, Reviewer), Vincent Feuillard (Statistical Expert, Renault, Examiner) and Claire Gormley (Professor, University College Dublin, Examiner).
- **Manuscript**: tel-03540827

MSc in Big Data

Rennes, France

Ensai (National School for Statistics and Data Analysis) — *Sept. 2016 - Oct. 2017*

- Dual degree program with Ensai engineering degree.
- Main topics: Statistics, Applied Mathematics, Computer Science.
- Training topics: assess, treat, and analyze massive amounts of heterogeneous data.
- Program taught entirely in English.

M.S. in Statistics (*Diplôme d'ingénieur*)

Rennes, France

Ensai (National School for Statistics and Data Analysis) — *Sept. 2014 - Oct. 2017*

- Training topics: Statistics, Econometrics and Computer Science.

Statistics of random processes

Aarhus, Danmark

Aarhus University — *Jan. 2016 - Jun. 2016*

- ERASMUS exchange.

CPGE MPSI/MP

Reims, France

Lycée Clemenceau — *Sep. 2011 - Jun. 2014*

Experience

Data Scientist on the EMPOWER project

Paris, France

IRMES (Insep) — *Jun. 2021 - Jan. 2022*

- Maximize the performance of elite female athletes by optimizing their training responses with adapted workloads in synergy with their physiology and menstrual cycle.
- Develop statistical models to determine athletes hormonal profiles and analyze responses to training and competition loads.
- Link: https://labos-recherche.insep.fr/fr/empower

Research Engineer

Guyancourt, France

Technocentre, Renault — *Jan. 2018 - Mar. 2021*

- Develop clustering methods for the analysis of autonomous vehicle Advanced Driver-Assistance Systems data.
- Create a Python package for Functional Data Analysis: FDApy.
- Use of Google Cloud Platform (Compute Engine and BigQuery) for the analysis of vehicle data.
- Software: Python and **R**.

Data Scientist (Intern)

Guyancourt, France

Technocentre, Renault — *Apr. 2017 - Oct. 2017*

- Implement a massive data analysis methodology for the validation of driving assistance systems.
- Software: Matlab.

Research Assistant

Fort Collins, USA

Colorado State University — *Jun. 2016 - Aug. 2016*

- Compare survey estimators for the *National Survey of College Graduation*.
- Software: **R**.

Skills

Programming Python, **R**, Bash

Tools Google Cloud Platform, Git, Continuous Integration

Languages French (mother tongue), English (TOEIC - 920 / 990)

Teaching Experience

Duration models

Ensai

M1 level — *Apr. 2021 - Jun 2021*

- Introduce duration models to ENSAI students (2nd year).
- Theoretical and practical (with **R**) sessions.

Linear regression

Ensai

M1 level — *Sep. 2019 - Nov. 2021*

- Introduce linear regression to ENSAI students (2nd year).
- Theoretical and practical (with **R**) sessions.

Scientific production

Articles

Clustering multivariate functional data using unsupervised binary trees.

CSDA

Golovkine S., Klutchnikoff N. & Patilea V. — *2021*

- Propose a model-based clustering algorithm for a general class of functional data for which the components could be curves or images.
- Apply to the analysis of vehicle trajectories on a German roundabout.
- DOI: 10.1016/j.csda.2021.107376

Adaptive optimal estimation of irregular mean and covariance functions

arXiv preprint

Golovkine S., Klutchnikoff N. & Patilea V. — *2021*

- Propose nonparametric estimators for the mean and covariance functions of functional data.
- Link: arXiv:2108.06507

FDApy: a Python package for functional data.

arXiv preprint

Golovkine S. — *2021*

- Explain how the package FDApy is implemented.
- Link: arXiv:2101.11003

Learning the smoothness of noisy curves with application to online curve estimation.

arXiv preprint

Golovkine S., Klutchnikoff N. & Patilea V. — *2020*

- Propose a simple estimator for the local regularity of the trajectories of a stochastic process.
- Real data sets illustrate the effectiveness of the new approach.
- Link: arXiv:2009.03652

Softwares

FDApy (`https://github.com/StevenGolovkine/FDApy`)

Python package

- Provide modules for the analysis of functional datasets.

denoisr (`https://github.com/StevenGolovkine/denoisr`)

R package

- Implement regularity estimation of a set of curves.

funestim (`https://github.com/StevenGolovkine/funestim`)

R package

- Implement mean and covariance estimation of a set of curves.

Conferences

With talk

Feb. 2022	**YSP,** 10th Young Statisticians and Probabilists day (invited)	*Virtual Conference*
Dec. 2021	**CMStatistics,** 14th conference on Computational and Methodological Statistics (invited)	*Virtual Conference*
Sep. 2021	**EYSM,** 22nd European Young Statisticians Meetings (invited)	*Virtual Conference*
Jun. 2021	**JDS,** 52nd Statistical Days	*Virtual Conference*
Mar. 2021	**Mathematics Seminars,** Hunter College (invited)	*Virtual Conference*
Dec. 2020	**CMStatistics,** 13th conference on Computational and Methodological Statistics (invited)	*Virtual Conference*
Nov. 2020	**StatMod2020,** Statistical Modeling with Applications (invited)	*Virtual Conference*
Jun. 2020	**JDS,** 52nd Statistical Days	*Conference Proceedings*
Jun. 2019	**JDS,** 51st Statistical Days	*Nancy, France*
Mar. 2019	**MASCOT-NUM,** Annual Conference	*Rueil-Malmaison, France*

Attended

Jun. 2019	**DS^3,** Data Science Summer School	*Saclay, France*