Erwan Scornet

Curriculum Vitae

Turing Building,
Ecole Polytechnique,
1 Rue Honoré d'Estienne d'Orves,
91120 Palaiseau France
✉ *erwan.scornet@polytechnique.edu*
https://erwanscornet.github.io/

Current Research

My research work focuses on statistical learning, in particular on non-parametric algorithms such as random forests. This class of algorithms exhibits good accuracy in practice but theoretical results do not entirely explain their good empirical performances. Therefore random forests theory is very exciting. I am also interested in the notion of explainable AI and in particular how to provide insights about how the algorithm works. I am also working on connections between random forests and neural networks, which are also non-parametric algorithms with tremendous empirical performance. More recently, I have started to work on missing value theory, trying to design algorithms that can easily handle this type of data.

Training

2009 **Admitted to École Normale Supérieure de Paris,** *by competitive examination based on mathematics and physics.*

2009-2010 **Bachelor in mathematics and physics,** *with honors (bien/très bien).*

2010-2011 **Master thesis on Metropolis-Hasting algorithm on Riemanian varieties.** *Several other research topics were spherical black holes, modelling growing interface, introduction to statistical learning.*

2011-2012 **Master degree on *Random modelling* from university Paris 7,** *with honors (très bien).*

Spring/Summer 2012 **Internship at Robert Debré Hospital (Paris) on glycemic variability for diabetic patients**.

2014 **Degree from École Normale Supérieure de Paris** *in mathematics (with a minor in physics).*

2012-2015 **PhD in statistics** *on "Learning with random forests"* supervised by Gérard Biau (LSTA - University Paris 6) et de Jean-Philippe Vert (Mines - Institut Curie)
Reviewers: Peter Bühlmann, Pierre Geurts.
Jury: Sylvain Arlot, Arnaud Guyader

March-April 2016 **Visiting Professor Zaid Harchaoui to the Courant Institute (New-York),**.

2016 - **Assistant Professor in Applied Mathematics,** *at École Polytechnique in Paris* (CMAP).

December 2020 **Habilitation à diriger des recherces** *Neural networks, interpretability, neural networks and missing values* Reviewers: Sylvain Arlot, Giles Hooker, Gabor Lugosi Jury: Florence D'Alché-Buc, Olga Klopp, Eric Moulines.

Teaching

2013-2015 **Probability and Statistics course** *for second year Bachelor students in economy*.

2013-2014 **Tutorial sessions in mathematics** *for second year Bachelor students in Paris Sciences et Lettres (PSL)*.

2012-2013 **Tutorial sessions in mathematics** *for first year Bachelor students in Paris Sciences et Lettres (PSL)*.

2016-2017 **Tutorial sessions in Machine Learning I** *for first year Graduate students in "Big data for business"* (X-HEC).

2016-2017 **Tutorial sessions in "The Art of Regression** *for third year engineering students*.

2016-2017 **Machine Learning Course** *for first year Graduate students in "Big data for business"* (X-HEC).

2016-2017 **Tutorial sessions for Machine Learning II Course** *for third year engineering students*.

2017-2018 **Machine Learning I course** *for first year Graduate students in "Data science for business"* (X-HEC).

2017-2018 **Learning Theory Course** *for third year engineering students*.

2017-2018 **Tutorial sessions for Machine Learning II Course** *for third year engineering students*.

2018-2021 **Deep Learning Course** *for second-year master students*.

2018-2021 **Optimization Course** *for second-year master students*.

2018-2021 **Learning Theory Course** *for third year engineering students*.

Besides teaching courses, I gave oral tests to Bachelor students for two years in some preparatory class for the "grandes école". From 2009 to 2014, I also participated in a mentoring program directed towards high school students from Paris suburb in order to encourage them to go to the university. I co-organized a week of intensive classes for high school students in August 2013 and August 2014.

Administrative work

2013, 2014 **Co-organizer of a week of intensive classes,** *for high school students* taking place at École Normale Supérieure.

2013 - 2015 **Co-organizer of the PhD students seminar,** *at LSTA, University Paris 6*.

2016 - 2017 **Co-organizer of the "Café des statistiques" weekly seminar,** *at École Polytechnique*.

2017, 2018 **Co-organizer of the one-day conference Young Statistician and Probabilists,** *at Institut Henri Poincaré*.

2017 - **In charge of data science third-year projects**.

2017 - 2018 **Co-head of the third year of Applied Mathematics** *at École Polytechnique*.

2017 - **Co-director of the Master of Science And Technology "Artificial Intelligence and Advanced Visual Computing",** *starting in September 2018 at École Polytechnique.*

2018 - **Co-director of the interships of first and second year** *in the "Artificial Intelligence and Advanced Visual Computing" training.*

2017, 2018 **Co-organizer of the Data Science Summer School (DS3),** *École Polytechnique.*

2017 - 2019 **Member of the scientific comittee of the DS3,** *École Polytechnique.*

June 2019 **Jury member of the PhD defense of Antonio Sutera,** *Liege university.*

2012 - **Reviewer** *for the Journal of Machine Learning Research, The Annals of Statistics, International Conference on Machine Learning 2015, Neural Information Processing Systems 2015.*

Programming skills

LaTeX, R, Python

Talks

December 2012 **PhD students seminar, LSTA** *Paris 6.*

March 2013 **Computational Biology team seminar** *Institut Curie.*

June 2013 **Journées de statistique** *Toulouse.*

December 2013 **NIPS, workshop MLCB** *Reno, Nevada.*

March 2014 **PhD students seminar, MAP5** *Paris 5.*

March 2014 **Poster at the colloquium "Digital: Big scale and complexity"** *Institut Mines-Télécom.*

April 2014 **"Random modelling and applications" seminar** *Caen.*

April 2014 **SMILE seminar** *ENS Ulm.*

June 2014 **Journées de statistique** *Rennes.*

January 2015 **Maths & companies week, organized by AMIES** *Paris.*

January 2015 **Young Statisticians and Probabilists conference** *Institut Henri Poincaré.*

February 2015 **Probability and statistics seminar** *Institut de Mathématiques et de Modélisation de Montpellier.*

June 2015 **Journées de statistique** *Lille.*

September 2015 **MODAL seminar** *INRIA Lille.*

October 2015 **Statistic seminar** *Université de Strasbourg.*

November 2015 **STA seminar** *Telecom ParisTech.*

January 2016 **Statistic seminar** *Université de Toulouse.*

June 2016 **Journées de statistique** *Montpellier.*

August 2016 **MAS conference** *Grenoble*.

September 2016 **Statistic seminar** *Compiègne*.

January 2017 **Statistic seminar** *Agro ParisTech*.

May 2017 **Statistic seminar** *MAP5*.

May 2017 **Statistic seminar** *École des ponts*.

July 2017 **Joint Statistical Meetings (JSM 2017)** *Baltimore*.

January 2018 **Invited speaker at a workshop on the Interface of Machine Learning and Statistical Inference** *Banff, Canada*.

March 2018 **Conference to prepare high school students to a conference by Yann LeCun** *Bourg La Reine*.

April 2018 **Statistic seminar** *Rennes*.

April 2018 **Statistic seminar** *Versailles*.

May 2018 **Statistic seminar** *Telecom ParisTech*.

January 2019 **Probability and Statistic seminar** *Lille university*.

January 2019 **Invited speaker at Journée Statistique / Apprentissage à Paris Saclay** *IHES*.

March 2019 **Invited speaker at BNP Cardiff conference** *Nanterre*.

March 2019 **Invited speaker at Sanofi conference** *Lyon*.

May 2019 **Statistic seminar** *Paris 7 university*.

October 2019 **Invited Speaker at Conference on Big Data and Machine Learning in Econometrics, Finance, and Statistics** *Chicago university*.

October 2019 **Invited speaker in a thematic class** *Puebla, Mexico*.

Award and distinctions

April 2016 **Grant from the Pierre Ledoux Foundation** *to fund a visit to the Courant Institute(New-York)* .

April 2016 **Recipient of the Jacques Neveu 2016 Prize,** *rewarding a PhD thesis in probability or statistics, received during the MAS conference* (Grenoble, août 2016).

Supervision - Internship

July 2013 **Co-supervision of the internship of Nelly Alandou, Clément Benesse et Pablo Le Henaff,** *first-year students in Paris Sciences et Lettres*.

Spring 2014 **Supervision of the internship of Arthur Pajot,** *third-year student at University Paris 6*.

Avril 2015 **Supervision of the internship of Charlotte Rougier,** *lycéenne en première S*.

2016 **Supervision of the internship of Jurriaan Parie,** *third-year student at University Paris 6*.

2016 - **Supervision of several third-year projects in data science,** *at École Polytechnique*.

Supervision - PhD

2016 - 2020 **Co-supervision with Stéphane Gaïffas (University Paris 7) of the PhD thesis of Jaouad Mourtada on *Contributions to statistical learning: density estimation, expert aggregation and random forests*.**

2018 - 2019 **Co-supervision with Julie Josse (Ecole Polytechnique), Gael Varoquaux (Inria) of the PhD thesis of Nicolas Prost.**

2018 - **Co-supervision with Gérard Biau (LPSM), Sebastien Da Veiga (Safran) of the PhD thesis of Clement Benard.**

2020 - **Co-supervision with Claire Boyer (LPSM) of the PhD thesis of Ludovic Arnould.**

2020 - **Co-supervision with Julie Josse (CMAP), Gaël Varoquaux (INRIA Saclay) of the PhD thesis of Bénédicte Colnet.**

Publications

2015 **Consistency of random forests,** *E. Scornet, G. Biau, and J.-P. Vert* The Annals of Statistics, Vol. 43, pp. 1716-1741.

2016 **On the asymptotics of random forests,** *E. Scornet* Journal of Multivariate Analysis, Vol. 146, pp. 72–83.

2016 **Random forests and kernel methods,** *E. Scornet* IEEE Transactions on Information Theory, Vol. 62, pp. 1485-1500.

2016 **A Random Forest Guided Tour,** *G. Biau and E. Scornet* TEST, Vol. 25, pp. 197-227, with Discussion.

2016 **Promenade en forêts aléatoires,** *E. Scornet* MATAPLI, Vol. 111.

2017 **Kernel multitask regression for toxicogenetics,** *E. Bernard, Y. Jiao, E. Scornet, V. Stoven, T. Walter and J.-P. Vert* Molecular Informatics, Vol. 36.

2017 **Universal consistency and minimax rates for online Mondrian Forest,** *J. Mourtada, S. Gaïffas, E. Scornet* NIPS 2017.

2017 **Tuning parameters in random forests,** *E. Scornet* ESAIM Procs, Vol. 60 pp. 144-162.

2018 **Impact of subsampling and tree depth on random forests,** *R. Duroux, E. Scornet* ESAIM: Probability and Statistics, Vol. 22, pp. 96-128.

2018 **Neural Random Forests,** *G. Biau, E. Scornet, J. Welbl,* Sankhya A, 1-40.

2020 **Minimax optimal rates for Mondrian trees and forests,** *J. Mourtada, S. Gaïffas, E. Scornet,* (Annals of Statistics 2020).

2020 **Linear predictor on linearly-generated data with missing values: non consistency and solutions,** *M. Le Morvan, N. Prost, J. Josse, E. Scornet., G. Varoquaux,* (AISTAT, 2020).

2019 **On the consistency of supervised learning with missing values,** *J Josse, N. Prost, E. Scornet, G. Varoquaux,* (in revision in JMLR).

2019 **AMF: Aggregated Mondrian Forests for Online Learning,** *J. Mourtada, S. Gaïffas, E. Scornet,* (in revision in JRSSB, 2020).

2020 **SIRUS: Stable and Interpretable RUle Set for Classification ,** *C. Bénard, G. Biau, S. Da Veiga, E. Scornet,* (accepted for publication in EJS).

2020 **Trees, forests, and impurity-based variable importance,** *E. Scornet,* (submitted).

2020 **Interpretable Random Forests via Rule Extraction,** *C. Bénard, G. Biau, S. Da Veiga, E. Scornet,* (submitted, 2020).

2020 **Neumann networks: differential programming for supervised learning with missing values,** *M. Le Morvan, J. Josse, T. Moreau, E. Scornet, G. Varoquaux* (Oral, Neurips 2020).

Hobbies

- From 2014 to 2019, Stage and lighting director of the musicals *Bloody Monday*, *Doré Mirador* and *Au bonheur des âmes*
- From 2018, amateur actor in the theater company *Dans de beaux drames*
- Tennis, badminton