Erwan Scornet

Curriculum Vitae



Research interests

My research work focuses on statistical learning theory, in particular non-parametric algorithms such as random forests. This class of algorithms exhibits good predictive performance that are not yet fully explained by theory. Therefore analyzing random forests from a theoretical perspective is both challenging and very exciting! I am also interested in the notion of explainable AI, i.e., providing insights about how algorithm predictions are built (variable importance, list of simple decision rules). I am also working on missing value theory, trying to design algorithms that can easily handle incomplete data and understand the impact of imputation on predictive performances. More recently, I started exploring the connections between random forests and neural networks, another kind of non-parametric algorithms with tremendous empirical performance. I also work on causal inference, studying the properties of causal measure in generalization settings (extending the finding from a clinical trial to a target population) and establishing asymptotic (or even non asymptotic) results on several estimators (G-formula, IPW, doubly-robust estimators...).

Keywords: Random forests, missing data, interpretability, causality, neural networks.

Training

- 2009-2013 **Degree from École Normale Supérieure de Paris** in mathematics (with a minor in 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).
 - 2012 **Internship at Robert Debré Hospital (Paris)** on glycemic variability for diabetic patients (6 months).
- 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 2016 Academic visit with Professor Zaid Harchaoui to the Courant Institute (New-York).

- 2016 2023 **Assistant Professor in Applied Mathematics,** at École Polytechnique in Paris (CMAP).
 - Dec. 2020 **Habilitation à diriger des recherches** Random forests, interpretability, neural networks and missing values.

 Reviewers: Sylvain Arlot, Giles Hooker, Gabor Lugosi
 Jury: Florence D'Alché-Buc, Olga Klopp, Eric Moulines.
 - 2023 ... Lecturer in Applied Mathematics, at Sorbonne Université in Paris (LPSM, SCAI).

Teaching

- 2012-2013 **Tutorial sessions in mathematics (TD, 24h)** for first-year bachelor students in Paris Sciences et Lettres (PSL).
- 2013-2014 **Tutorial sessions in mathematics (TD, 24h)** for second-year bachelor students in Paris Sciences et Lettres (PSL).
- 2013-2015 **Probability and Statistics (New course, 48h)** for second-year bachelor students in economy.
- 2016-2019 **Tutorial sessions in Machine Learning I (TD/TP, 18h)** for first-year graduate students in "Data science for business" (X-HEC).
- 2016-2017 **Tutorial sessions in "The Art of Regression" (TD, 36h)** for third-year engineering students.
- 2016-2018 **Tutorial sessions in Machine Learning II (TD/TP, 18h)** for third-year engineering students.
- 2016-2018 **Tutorial sessions in "Statistics in Action" (TD/TP, 18h)** for third-year engineering students.
- 2017-2019 **Machine Learning Course (Course, 9h)** for first-year graduate students in "Data science for business" (X-HEC).
- 2018-2023 **Deep Learning Course (New course, 21h)** for second-year master students.
- 2019-2020 Tutorial sessions in Deep Learning (TP, 21h) for second-year master students.
- 2019-2023 **Deep Learning and Optimization Course (lectures and lab sessions)** for first-year graduate students in "Data science for business" (X-HEC).
- 2018-2023 **Learning Theory Course (New course, 12h)** for third year engineering students.
- 2022-2024 MOOC Machine and Deep Learning, AI Certificate, IP Paris.
- 2023-2026 **Data science Course and Tutorials (40h)** for non-specialized second-year bachelor students.
- 2023-2026 Course and tutorial on tree-based methods and deep learning (22h) for M2 students.
- 2023-2026 Course and tutorial on machine learning and CNN (one week) for non-specialized students (medicine).

Besides teaching courses, I gave oral examination training to Bachelor students for two years in some preparatory class for the "grandes écoles". From 2009 to 2014, I also participated in a monthly mentoring program named TalENS directed towards high school students from Paris suburb in order to encourage them to keep studying beyond their 'baccalauréat'. In particular, I co-organized a week

of intensive classes for high school students in August 2013 and August 2014.

Past responsibilities

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

Current responsibilities

- 2012 **Reviewer** for the Journal of Machine Learning Research, The Annals of Statistics, Electronic Journal of Statistics, International Conference on Machine Learning, Neural Information Processing Systems and others.
- 2017 2023 In charge of data science third-year projects.
- 2017 2023 Co-director of the Master (MScT) "Artificial Intelligence and Advanced Visual Computing", created in September 2018 at École Polytechnique.
- 2018 2023 **Co-director of the internships of first and second year** in the "Artificial Intelligence and Advanced Visual Computing" training.
- 2020 2023 **Scientific responsible of the Artificial intelligence program** of Labex Mathématiques Hadamard (LMH) with Frédéric Chazal.
 - 2023 ... **Co-director of the minor degree in Data Science and Artificial Intelligence** for second and third-year Bachelor students at Sorbonne University.

Jury

- April 2025 Jury member of Inria committee recruitment for CR positions, Paris center.
- November **Jury member of the PhD defense of Julien Demange-Chryst,** "Nouvelles 2024 méthodes d'estimation par échantillonnage préférentiel pour l'analyse de sensibilité et l'estimation d'évènements rares", Toulouse III University, France.
- October 2024 **Reviewer of the PhD defense of Virgile Foy,** "Apprentissage de représentations sur des surfaces 3D et analyse de nuages de points, application à la conception aéronautique assistée par ordinateur", Toulouse III University, France.
 - July 2024 **Jury member of the PhD defense of Amin Dhaou,** "Interpretable and Causal Analysis for Multivariate Time Series", École polytechnique, France.
 - June 2024 President of the jury for the PhD Defence of Ahmad Chamma, "Statistical Interpretation of High-Dimensional Complex Prediction Models for Biomedical Data", Inria Saclay, France.

- February Reviewer of the PhD defense of Florian Lalande, "Planetary Systems Insights
 - 2024 through Numerical Data Imputation Algorithms and Machine Learning", Okinawa Institute of Science and Technology, Japan.
- February Reviewer of the PhD defense of Jean-Samuel Leboeuf, "On the generalization
 - 2023 properties of VC classes and application to decision trees", Laval University, Canada.
- December Jury member of the PhD defense of Clément Bénesse, "On the links between
 - 2022 Global Sensitivity Analysis and Algorithmic Fairness for eXplainable and Fair Machine Learning", Toulouse III University, France.
- December Jury member of the PhD defense of Baptiste Kerleguer, "Multi-fidelity surrogate
 - 2022 modeling adapted to functional outputs for uncertainty quantification of complex models", École polytechnique, France.
- June 2019 **Jury member and reviewer of the PhD defense of Antonio Sutera,** "Importance measures derived from random forests: characterisation and extension", Liege University, Belgium.
- June 2019 **Jury member and reviewer of the PhD defense of Irving Gomez Mendez,** *entitled "Random forests and autoencoders with missing data",* CIMAT, Mexico.
- June 2021 Member of the selection committee for an assistant professor position, Université Gustave Eiffel.
- June 2022 Member of the selection committee for an assistant professor position, *Université de Bordeaux*.

Talks

- Dec. 2012 PhD students seminar, LSTA, Paris 6.
- March 2013 Computational Biology team seminar, Institut Curie.
 - June 2013 Journées de statistique, SFDS, Toulouse.
- Dec. 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", IMT.
 - April 2014 "Random modelling and applications" seminar, Caen.
 - April 2014 SMILE seminar, ENS Ulm.
 - June 2014 Journées de statistique, SFDS, Rennes.
- January 2015 Maths & companies week, organized by AMIES, Paris.
- January 2015 Young Statisticians and Probabilists conference, Institut Henri Poincaré.
 - Feb. 2015 **Probability and statistics seminar,** *Institut de Mathématiques et de Modélisation de Montpellier.*
 - June 2015 Journées de statistique, SFDS, Lille.
 - Sept. 2015 MODAL seminar, INRIA Lille.
- October 2015 Statistic seminar, Université de Strasbourg.
 - Nov. 2015 STA seminar, Telecom ParisTech.
- January 2016 Statistic seminar, Université de Toulouse.

- June 2016 Journées de statistique, SFDS, Montpellier.
- August 2016 MAS conference, Grenoble.
- Sept. 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.
- January 2021 **Statistic seminar,** Angers.
 - March 2021 **Statistic seminar for M2 students,** *University Paris-Saclay.*
 - May 2021 Statistical workshop, Amsterdam School of Economics.
 - July 2021 ENBIS Workshop: Interpretability for Industry 4.0, Naples.
- January 2022 Statistical seminar for M2 students, Ecole Polytechnique.
 - Feb. 2022 **Statistical seminar,** Berlin, Weierstrass Institute.
 - Feb. 2022 Statistical seminar, Orsav.
 - March 2022 Statistical seminar, Vannes.
 - March 2022 Statistical seminar, Montpellier.
 - April 2022 **Statistical seminar,** Centre Borelli, ENS Paris-Saclay.
 - June 2022 **Journées de statistique, SFDS,** *Lyon.*
 - Sept. 2022 Statistical and Optimization seminar, Toulouse.
 - Nov. 2022 **Econometrics seminar,** Cambridge.
 - Dec. 2022 International Conference on Statistics and Data Science, Florence.
 - Feb. 2023 **Econometrics seminar,** *Amsterdam University*.
 - Feb. 2023 Data Science seminar, London School of Economics.
 - April 2023 Stat'Learn conference, Montpellier.

- June 2023 Quarter on Causality, conference, Institut Pascal, Paris Saclay.
- Dec. 2023 ICSDS Conference, Lisbonne.
- January 2024 Round table on Al challenges, Artefact research center opening, Paris.
 - May 2024 SIERRA Team, INRIA Paris.
 - June 2024 Workshop "Mathematical foundation of Al", organized by SCAI, Sorbonne University Paris.
 - Dec. 2024 SAMM seminar, University Paris 1.
 - Dec. 2024 Seminar of students in statistics, Orsay Paris.
- January 2024 MAP5 seminar, University Paris Cité Paris.
 - Feb. 2025 Statistics and Optimisation seminar, IMT Toulouse.
 - Feb. 2025 PreMedical seminar, Inria Montpellier.
 - March 2025 Statistical Parisian seminar, IHP Paris.
 - April 2025 Statistics and Data Science conference Milan.

Grant and Awards

- April 2016 **Grant from the Pierre Ledoux Foundation** *to fund a visit to the Courant Institute,* New-York.
- August 2016 Recipient of the Jacques Neveu 2016 Prize, rewarding a PhD thesis in probability or statistics, received during the MAS conference, Grenoble.
 - June 2018 Missing DatalA project funded by DatalA institute.
 - June 2020 PhD thesis funded by SCAI (Sorbonne Center for Artificial Intelligence) on neural networks.
 - June 2021 **Recipient of an Emergence Project,** provided by Paris city to develop a topic in machine learning (150.000 euros).

Supervision - Internships

- 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.
 - 2020 **Co-supervision of the M2 internship of Ludovic Arnould** with Claire Boyer (LPSM).
 - 2021 **Co-supervision of the post-doctoral research of Marine Le Morvan** with Julie Josse (INRIA Montpellier) and Gaël Varoquaux (INRIA Saclay).

- 2021 **Co-supervision of the M2 internship of Alexis Ayme** with Claire Boyer (LPSM) and Aymeric Dieuleveut (CMAP).
- 2022 Co-supervision of the M2 internship of Patrick Lutz with Claire Boyer (LPSM).
- 2022 **Co-supervision of the M2 internship of Khadim Sene** *with Alexandre Py-Renaudie (IPP, IPVF) and Jean-François Guillemoles (IPP, IPVF).*

Supervision - PhD

- 2016 2020 **Co-supervision of the PhD thesis of Jaouad Mourtada** *entitled Contributions to statistical learning: density estimation, expert aggregation and random forests* with Stéphane Gaïffas (University Paris 7).
- 2018 2019 **Co-supervision of the PhD thesis of Nicolas Prost** with Julie Josse (Ecole Polytechnique) and Gael Varoquaux (Inria).
- 2018 2021 **Co-supervision of the PhD thesis (CIFRE) of Clement Benard** *entitled Forêts aléatoires et interprétabilité des algorithmes d'apprentissage* with Gérard Biau (LPSM) and Sebastien Da Veiga (Safran).
- 2020 2023 Co-supervision of the PhD thesis of Ludovic Arnould with Claire Boyer (LPSM).
- 2020 2023 **Co-supervision of the PhD thesis of Bénédicte Colnet** with Julie Josse (INRIA Montpellier) and Gaël Varoquaux (INRIA Saclay).
- 2021 2024 **Co-supervision of the PhD thesis of Alexis Ayme** with Claire Boyer (LPSM) and Aymeric Dieuleveut (CMAP).
 - 2023 ... **Co-supervision of the PhD thesis of Ahmed Boughdiri** with Julie Josse (Inria Montpellier).
 - 2024 ... Co-supervision of the Cifre PhD thesis of Abdoulaye Sakho with Emmanuel Malherbe (Artefact)).
 - 2025 ... Co-supervision of the Cifre (Withings) PhD thesis of Agathe Chabassier with Julie Josse (Inria Montpellier).
 - 2025 ... **Co-supervision of the PhD thesis of Charif El Gataa** with Christophe Denis (Paris 1) and Julie Josse (Inria Montpellier).
 - 2025 ... **Co-supervision of the PostDoc Simon Roburin** with Rafaël Pinot (Sorbonne Université).

Book

2022 Interpretability for Industry 4.0 : Statistical and Machine Learning Approaches, B. looss, R. Kenett, P. Secchi, B.M. Colosimo, F. Centofanti, C. Bénard, S. Da Veiga, E. Scornet, S. N. Wood, Y. Goude, M. Fasiolo Editors: A. Lepore, B. Palumbo, J.-M. Poggi, Springer.

Published papers

- 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 **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,* The Annals of Statistics, 48(4), 2253-2276.
- 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 Neumann networks: differential programming for supervised learning with missing values, M. Le Morvan, J. Josse, T. Moreau, E. Scornet, G. Varoquaux Oral, NeurlPS.
- 2021 SIRUS: Stable and Interpretable RUle Set for Classification , C. Bénard, G. Biau, S. Da Veiga, E. Scornet, Electronic Journal of Statistics 2021, Vol. 15, pp. 427-505.
- 2021 Interpretable Random Forests via Rule Extraction, C. Bénard, G. Biau, S. Da Veiga, E. Scornet, AISTAT.
- 2021 **AMF: Aggregated Mondrian Forests for Online Learning,** *J. Mourtada, S. Gaïffas, E. Scornet,* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 83(3), 505-533.
- 2021 Analyzing the tree-layer structure of Deep Forests, L. Arnould, C. Boyer, E. Scornet, ICML.
- 2021 What's a good imputation to predict with missing values? M. Le Morvan, J. Josse, E. Scornet, G. Varoquaux Oral, NeurIPS.
- 2021 **Trees, forests, and impurity-based variable importance,** *E. Scornet,* Annales de l'Institut Henri Poincaré.
- 2022 **SHAFF:** Fast and consistent **SHApley** eFfect estimates via random Forests, *C. Bénard, G. Biau, S. Da Veiga, E. Scornet,* AISTAT.
- 2022 MDA for random forests: inconsistency, and a practical solution via the Sobol-MDA, C. Bénard, S. Da Veiga, E. Scornet, Biometrika.

- Near-optimal rate of consistency for linear models with missing values, *A. Ayme, C. Boyer, A. Dieuleveut, E. Scornet* ICML.
- 2022 **Generalizing a causal effect: sensitivity analysis and missing covariates,** *B. Colnet, J. Josse, E. Scornet, G. Varoquaux* Journal of Causal Inference.
- 2023 **Is interpolation benign for random forests?,** *L. Arnould, C. Boyer, E. Scornet* AISTAT.
- 2023 **Sparse tree-based initialization for neural networks,** *P. Lutz, L. Arnould, C. Boyer, E. Scornet.* ICLR.
- 2023 Naive imputation implicitly regularizes high-dimensional linear models, *A. Ayme, C. Boyer, A. Dieuleveut, E. Scornet.* ICML.
- 2024 (first submission in 2019) On the consistency of supervised learning with missing values, J. Josse, J.M. Chen, N. Prost, E. Scornet, G. Varoquaux Statistical Papers.
- 2024 Reweighting the RCT for generalization: finite sample error and variable selection, B. Colnet, J. Josse, G. Varoquaux, E. Scornet JRSS-A.
- 2024 Random features models: a way to study the success of naive imputation, A. Ayme, C. Boyer, A. Dieuleveut, E. Scornet. ICML.
- 2025 **A primer on linear classification with missing data,** A.D. Reyero Lobo, A. Ayme, C. Boyer, E. Scornet AISTAT.
- 2025 Quantifying Treatment Effects: Estimating Risk Ratios via Observational Studies, A. Boughdiri, J. Josse, E. Scornet ICML.
- 2025 Harnessing Mixed Features for Imbalance Data Oversampling: Application to Bank Customers Scoring, A. Sakho, E. Malherbe, C.-E. Gauthier, E. Scornet ECML-PKDD, Applied Data Science track.

Submitted papers

- 2023 Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize?, B. Colnet, J. Josse, G. Varoquaux, E. Scornet.
- 2024 **Do we need rebalancing strategies? A theoretical and empirical study around SMOTE and its variants.,** *A. Sakho, E. Malherbe, E. Scornet.*
- 2024 What Is a Good Imputation Under MAR Missingness?, J. Naf, J. Josse, E. Scornet.
- 2025 **Theory of Random Forests: A Review,** E. Scornet, G. Hooker.
- 2025 A Unified Framework for the Transportability of Population-Level Causal Measures, A. Boughdiri, C. Berenfeld, J. Josse, E. Scornet.
- 2025 Causal Meta-Analysis: Rethinking the Foundations of Evidence-Based Medicine, C. Berenfeld, A. Boughdiri, B. Colnet, W. A. C. van Amsterdam, A. Bellet, R. Khellaf, E. Scornet, J. Josse.
- 2025 Asymptotic Normality of Infinite Centered Random Forests Application to Imbalanced Classification, M. Mayala, E. Scornet, C. Tillier, O. Wintenberger.

- 2025 When Pattern-by-Pattern Works: Theoretical and Empirical Insights for Logistic Models with Missing Values, C. Muller, J. Josse, E. Scornet.
- 2025 How to rank imputation methods? J. Näf, K. Grzesiak, E. Scornet.

Hobbies

- From 2014 to 2019, Stage and lighting director of the musicals *Bloody Monday*, *Doré Mirador* and *Au bonheur des âmes*, original creation and production with a troupe of 40 artists.
- From 2018, amateur actor in the theater company *Dans de beaux drames* in *Le repas des fauves* (2019) and *Le porteur d'histoire* (2021).
- Chess, Tennis, Badminton, application of probability theory to understand board games ("Mon premier vergé").