Julie Josse

Nationality: French
Date of birth: 10th November 1983
E-mail: julie.josse@inria.fr
Web: http://juliejosse.com

Position
Fev 2023 - **Apple Research (1 day/week)**
June 2022 - **Head of PreMeDICaL team (Inria-Inserm team)**
Sept. 2020 - **Senior Researcher, Inria**

Working Address: Inria, Antenne de Montpellier,
860 rue Saint Priest, 34000 Montpellier

Education - previous positions

- Dec. 2020 - Nov. 2022 Scientific collaborator and Professor, École Polytechnique

- Sept. 2016 - Aug. 2020. Professor of Statistics at **École Polytechnique** (Institut Polytechnique de Paris), centre de mathématiques appliquées (CMAP) department
  Member of the Inria team Xpop

- Apr. 2019 - Sept. 2020. Visiting researcher (2 days a week). **Google Brain** IA, Paris

- Aug. 2016. **Habilitation à diriger des recherches mathématique** Paris-Sud University. *"Contribution to missing values and principal components methods"*. **Committee**. Reviewers: G. Celeux (Inria), J. Friedman (Stanford Univ.), P. Hoff (Washington Univ.); members: A. Ruiz-Gazen (Toulouse 1 Univ.), P. Massart (Paris-Sud Univ.), E. Moulines (École Polytechnique), F. Murtagh (London Univ.), D. Paindaveine (Bruxelles Univ.)

- 2016. Research associate (délégation) **Inria** Select team, Paris-Sud University, Orsay

- 2011 - 2015. Associate professor, **Agrocampus Ouest** (grande école d'ingénieurs, Ministry of agriculture), Statistics department, Rennes

- 2007–2010: **PhD thesis in Statistics defended 22 October 2010.** Agrocampus, Rennes. *"Gestion des données manquantes en analyse exploratoire des données"*. Advisors: J. Pagès (Agrocampus) & F. Husson (Agrocampus). **Committee**: M. Greenacre (Pompeu Fabra Univ.), H.A.L. Kiers (Groeningen Univ., reviewer), G. Govaert (Compiegne Univ., reviewer), P. Kroonenberg (Leiden Univ.), A. Morin (Rennes 1, Univ.)

  Award: best doctoral PhD thesis in applied statistics by the French Statistical Society (Prix Marie-Jeanne Laurent-Duhamel)

- 2006 - 2010. Engineer (Ingénieur d'études) **Agrocampus**, Rennes

- 2004 - 2006. Master degree Statistic *(summa cum laude)* . **Rennes II Univ.**

Publications

Articles

1. Colnet, B., Josse, J., Varoquaux, G. and Scornet, E. (2023). Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize? *Submitted*

2. Zaffran, M., Josse, J., Dieuleveut, A., Romano, Y. Conformal prediction with missing values. *ICML2023*

3. Zhao, P., Josse, J. Yang, S. (2023). Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data. *Submitted.*

4. Blet et al. (2022). Association between in-ICU red blood cells transfusion and one-year mortality in ICU survivors.
Critical Care.

5. Colnet, B., Josse, J., Varoquaux, G. and Scornet, E. (2022). Reweighting the RCT for generalization: finite sample analysis and variable selection. *Submitted.*

6. Gauss et al. (2022). Is Early Norepinephrine Associated with 24-hour Mortality of Blunt Trauma Patients in Haemorrhagic Shock? An International Cohort Study.
JAMA Network Open.

7. Colnet, B., Josse, J., Varoquaux, G. and Scornet, E. (2022). Generalizing a causal effect: sensitivity analysis and missing covariates.
Journal of Causal Inference.

8. Garaix et al. (2022). Decision-making tools for healthcare structures in times of pandemic.
Editorial Anaesthesia Critical Care & Pain Medicine.

9. Zaffran et al. (2022). Adaptive conformal prediction for time series.
ICML2022.

10. Perez-Lebel et al. (2022). Benchmarking missing-values approaches for predictive models on health databases.
GigaScience.

11. Sportisse et al. (2021). Model-based Clustering with Missing Not At Random Data.

12. Le Morvan, M., Josse, J., Scornet., E. and Varoquaux, G. (2021). What's a good imputation to predict with missing values?
Neurips2021, Oral.

13. Colnet et al. (2021-23). Causal inference methods for combining randomized trials and observational data: a review.
Statistical Science.

14. Mayer, I., Josse, J. (2021). Transporting treatment effects with incomplete attributes.
Biometrics.

15. Moyer et al. (2020). Trauma reloaded: Trauma registry in the era of data science Anaesthesia.
Critical Care & Pain Medicine.

16. Le Morvan et al. (2020). NeuMiss networks: networks: differential programming for supervised learning with missing values.
Neurips2020. Oral

17. Sbidian et al. (2020). Hydroxychloroquine with or without azithromycin and in-hospital mortality or discharge in patients hospitalized for COVID-19 infection: a cohort study of 4,642 in-patients in France.

18. Consortium ICUBAM. (2020). ICU Bed Availability Monitoring and analysis in the Grand Est région of France during the COVID-19 epidemic.
Statistiques et Société.

19. Sportisse, A., Josse, J. , Boyer, C. and Dieuleveut, A. (2020). Debiasing Stochastic Gradient Descent to handle missing values.
Neurips2020.

20. Sportisse, A., Boyer, C. and Josse, J. (2020). Estimation and imputation in Probabilistic Principal Component Analysis with Missing Not At Random data.
Neurips2020.

21. Le Morvan, M., N. Prost, J. Josse, E. Scornet. and G. Varoquaux. (2020). Linear predictor on linearly generated data with missing values: nonconsistency and solutions.
AISTAT2020.

22. Mayer, I., Wager, S., Sverdr, E., Moyer, J.D., Gauss, T. and Josse, J. (2020). Doubly robust treatment effect estimation with incomplete confounders.
Annals Of Applied Statistics.

23. Muzellec, B. Josse, J. , Boyer, C. and Cuturi, M. (2020). Missing Data Imputation using Optimal Transport.
International Conference on Machine Learning ICML2020.

24. Jiang, W., Bogdan, M., Josse, J., Miasojedow, B. and Rockova, V. (2020). Adaptive Bayesian SLOPE, High dimensional Model Selection with Missing Values.
Journal of Computational and Graphical Statistics.

25. Josse, J., Mayer, I. and Vert, J.P. (2020). MissDeepCausal: causal inference from incomplete data using deep latent variable models.

26. Descloux, P. Boyer, C., Josse, J., Sportisse, A. and Sardy, S. (2022). Robust Lasso-Zero for sparse corruption and model selection with missing covariates.
Scandinavian Journal of Statistics.

27. Mayer, I., Sportisse, A., Vialaneix, N., Tierney, N. and Josse, J. (2022). R-miss-tastic: a unified platform for missing values methods and workfows.
R journal.

28. Josse, J., Prost, N., Scornet, E. and Varoquaux, G. On the consistency of supervised learning with missing values. *Submitted.*

29. Sportissse, A., Boyer, C. and Josse, J. (2019). Imputation and low-rank estimation with Missing Non At Random data.
Statistics & Computing.

30. Robin, G., Klopp, O., Josse, J., Moulines, E. and Tibshirani, Rob. (2019). Main effects and interactions in mixed and incomplete data frames.
Journal of American Statistical Association.

31. Jiang, W., Lavielle, M. and Josse, J. (2018). Logistic Regression with Missing Covariates: Parameter Estimation, Model Selection and Prediction within a Joint-Modeling Framework.
Computational Statistics & Data Analysis.

32. Robin, G., Hoi To Wai, Josse, J., Klopp, O. and Moulines, E. (2018). Low-rank interactions and sparse additive effects model for large data frames.
NeurIPS.

33. Josse, J., Husson, F., Robin, G. and Balasubramanian, N. (2018). Imputation of mixed data with multilevel SVD.
Journal of Computational and Graphical Statistics.

34. Mozharovskyi, P., Husson, F. and Josse, J. (2018). Nonparametric imputation by data depth.
Journal of the American Statistical Association.

35. Hamada et al. (2018). Effect of Fibrinogen administration on early mortality in traumatic haemorrhagic shock: a propensity score analysis.
Journal of Trauma and Acute surgery.

36. Seijo-Pardo, B., Alonso-Betanzos, A., P. Bennett, K. Bolón-Canedo, Josse, J., Saeed, M., Guyon, I. (2018). Feature selection in the presence of missing data. *Neurocomputing, ESANN.*

37. Josse, J. and Reiter, J. (2018). Introduction to the Special Section on Missing Data.
Journal of Statistical Sciences.

38. Robin, G, Sardy, S., Moulines, E. and Josse, J. (2018). Low rank log-linear models for contingency tables.
Journal of Multivariate Analysis.

39. Josse, J. and Holmes, S. (2017). Discussion of 50 Years of Data Science.
Journal of Computational and Graphical Statistics.

40. Bollmann, S., Cook, Di. Dumas, J., Fox, J., Josse, J., Keyes, O. Strobl, C., Turner, H. & Debelak, R. (2017). A First Survey on the Diversity of the R Community.
R journal.

41. Josse, J., Marin, J.M. and Robert, C.P. Some discussions on the Read Paper "Beyond subjective and objective in statistics" by A. Gelman and C. Hennig.

42. Sobczyk, P., Bogdan, M. and Josse, J. (2017). Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood.
    Journal of Computational and Graphical Statistics.

43. Foulley, J.L., Celeux, G. and Josse, J. Empirical Bayes approaches to PageRank type algorithms for rating scientific journals. *Preprint.*

44. Fithian, W. and Josse, J. (2016). Multiple Correspondence Analysis & the Multilogit Bilinear Model.
    Journal of Multivariate Analysis.

45. Husson, F., Josse, J. and Saporta, G. (2016). Jan de Leeuw and the French school of data analysis.
    Journal of Statistical Software.

46. Groenen, P. and Josse, J. (2016). Multinomial multiple correspondence analysis. *Preprint.*

47. Josse, J., Sardy, S. and Wager, S. (2016). denoiseR a package for low rank-matrix estimation. *accepted in Journal of Statistical Software.*

48. Fujii, H., Josse, J., Tanioka, M., Miyachi, Y., Husson, F. and Ono M. (2016). Regulatory T cells in melanoma revisited by a computational clustering of FOXP3+ T cell subpopulations.
    Journal of Immunology.

49. Josse, J. and Wager, S. (2016). Stable Autoencoding: A Flexible Framework for Regularized Low-Rank Matrix Estimation.
    Journal of Machine Learning Research.

50. Audigier, V., Husson, F. and Josse, J. (2016) MIMCA: Multiple imputation for categorical variables with multiple correspondence analysis.
    Statistics and Computing.

51. Josse, J. and Holmes, S (2015). Tests of independence and Beyond.
    Statistics Survey.

52. Audigier, V., Josse, J. and Husson, F. (2015). Multiple imputation for continuous variable using Bayesian PCA.
    Journal of Statistical Computation and Simulation.

53. Josse, J. and Sardy, S. (2015). Adaptive Shrinkage of singular values.
    Statistics and Computing.

54. Josse, J. and Husson, F. (2015). missMDA a package to handle missing values in and with multivariate data analysis methods.
    Journal of Statistical Software.

55. Audigier, V., Husson, F. and Josse, J. (2014). A principal components method to impute mixed data.
    Advances in Data Analysis and Classification.

56. Josse, J., Wager, S. and Husson, F. (2014). Confidence areas for fixed-effects PCA.
    Journal of Computational and Graphical Statistics.

57. Dray, S and Josse, J. (2014). Principal component analysis with missing values: a comparative survey of methods.
    Plant Ecology.

58. Josse, J., van Eeuwijk, F., Piepho and H-P, Denis, J.B. (2014). Another look at Bayesian analysis of AMMI models for genotype-environment data.
    Journal of Agricultural, Biological, and Environmental Statistics.

59. Verbanck, M., Josse, J. and Husson, F. (2015). Regularized PCA to denoise and visualise data.
    Statistics and Computing.

60. Josse, J., Timmerman, M.E. and Kiers, H.A.L. (2013). Missing values in multi-level simultaneous component analysis.
    Chemometrics and Intelligent Laboratory Systems.

61. Josse, J., and Husson, F. (2013). Handling missing values in Multiple Factor Analysis.
    Food Quality and Preferences.

62. Josse, J., and Husson, F. (2013). Handling missing values in multivariate exploratory data analysis.
    Journal of the French Statistical Society (SFdS). (Paper written for the best Ph.D doctoral thesis prize delivered by the French Statistical Society).

63. Josse, J., Chavent, M., Liquet, B. and Husson, F. (2012). Handling missing values with Regularized Iterative Multiple Correspondence Analysis.
    Journal of Classification.

64. Denis, J.B. and Josse, J. (2012). Infering biadditive models within the Bayesian paradigm. *Technical report*: R2012. *http://prodinra.inra.fr/record/50125.*

65. Josse, J. and Husson, F. (2011). Selecting the number of components in PCA using cross-validation approximations.
    Computational Statistics and Data Analysis.

66. Josse, J., Pagès, J. and Husson, F. (2011). Multiple Imputation in PCA.
    Advances in Data Analysis and Classification.

67. Husson, F., Josse, J. and Pagès, J. (2010). Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? *Technical report.*

68. Josse, J., Pagès, J. and Husson, F. (2009). Données manquantes en analyse en composantes principales.
    Journal de la Société Française de Statitique.

69. Josse, J., Pagès, J. and Husson, F. (2008). Testing the significance of the RV coeffcient.
    Computational Statistics and Data Analysis.

70. Lé, S., Josse, J. and Husson, F. (2008). FactoMineR: an R package for multivariate analysis.
    Journal of Statistical Software.

Books

- Husson et al. (2018) R pour les statistiques et la data-science, Presses Universitaires de Rennes. `https://r-stat-sc-donnees.github.io/`. 4000 books sold.

- Cornillon et al. (2012). R for Statistics, Chapman & Hall. 320 pages. 1000 books sold. `http://www.crcpress.com/product/isbn/9781439881453`.

- Cornillon et al. (2008, 2010, 2012). Statistiques avec R, Presses Universitaires de Rennes. 300 pages. More than 10000 books sold.

Chapter in books

- Husson, F. and Josse, J. (2014). Multiple Correspondence Analysis. In: The Visualization and Verbalization of Data. Chapman & Hall/CRC. Edited by Blasius, J. and Greenacre, M.

Grants and Awards

- Advisory Board of HORIZON EUROPE (HORIZON-HLTH-2022-TOOL-11-02), more-europa.

- PEPR "Programme et Équipements Prioritaires de Recherche" in digital health, French National Research Agency (ANR). Co-lead of the workpackage *Enriching clinical trials with multi-sources and multidimensional auxiliary data.* (Start Sept 2023).

- PI of the Educational grant by (Allergologisk Laboratorium Kobenhavn) ALK, *Combining RCT and observational data.* (2022), (30K€).

- Member of Projet de Recherche Collaborative International France - Germany (ANR/DFG) (2021), *ASCAI.* PI: Nicolas Verzelen and Alexandra Carpentier. (275K€).

- PI Chaire Muse, French National Research Agency (ANR) (2021), *Premedical: precision medicine by causal learning.* (300K€).

- Member of Project Sorbonne Center for AI (SCAI) (2020), *Apprentissage Statistique pour les hopitaux.* PI: Claire Boyer. (70K€).

- Member of Independent Research Fund Denmark Grant (2019), *Missing data imputation using deep generative models.* PI: Jes Frellsen. (400K€).

- co-PI of Capgemini Invent support (2019 -), "Mécénat de compétences", *Traumatrix project* 10 persons allocated (data managers, data-scientists, project managers) to create a decision tool to improve the care of trauma patients.

- co-PI of DataIA institute grant (2018), *Missing data in the big data era* with Gael Varoquaux, (250K€).

- PI of R consortium grant (2018), *A unified platform for missing values methods and workflows.* (10K€).

- PI of PEPS grant (2017), *Handling missing data with heterogeneous data - modeling severe polytrauma patients.* (50K€).

- AgreenSkills grant (Marie-Curie FP7 COFUND People Program) (2013) for mobility research (18 months at Stanford University). It is an international program co-funded by the European Union designed to increase the research potential of promising researchers.

- Ph.D. dissertation award (Marie-Jeanne Laurent-Duhamel - French Statistical Society), (2012) it rewards the quality of a young, french statistician's doctoral thesis. In 2012, it distinguishes a notable contribution to applied statistical research.

Talks in conferences, workshops, seminars

Keynotes talks in conferences and workshops

- 2nd Global Symposium of Research Methodology Innovation in Trauma and Emergency Care, May 2024, Columbus Ohio.

- Journées de biostatistique Nov. 2023, Toulouse.

- Celebrating causal inference in medicine and public health, Oct. 2022, Ghent University, Belgium (before the Rousseeuw prize)

- Exposome Symposium, Oct. 2022, Montpellier, France

- AutoML conference (with ICML conference), Online, July 2022, Baltimore, USA

- Workshop, Emerging Challenges for Statistics and Data Sciences: Complex Data with Missingness, Measurement Errors, and High Dimensionality, Online, May 2022, University of British Columbia Okanagan (UBCO), Canada

- Open Data Science Conference, East Virtual, April, 2022

- DAG Stat. Keynote in Section Compstat, Online, March, 2022, Hambourg, Germany

- European Statistical Forum Application of Causal Inference in Drug Development, Online, Nov. 2021

- Optimization, Statistics, and Theoretical Machine Learning, Online, Sept., 2020, Princeton, USA

- Workshop in the international society for clinical biostatistic, Online, Aug. 2020, Krakow, Poland

- Workshop Mathematical Methods of Modern Statistics, Online, June 2020, CIRM, Marseille, France

- International Federation of Classification Societies, Aug. 2019, Thessaloniki, Greece

- useR!2019, July 2019, Toulouse, France

- R in Official Statistics, May 2019, Bucharest, Romania

- French Statistical Society, June 2019, Nancy, France

- Statlearn 2019, Grenoble, France

- R à Québec, July 2018, Montreal, Canada

- Insurance, Actuarial Science, Data and Models, June 2018, Paris, France

- Older keynotes talks: Finnish Workshop on Invariant Coordinate Selection. Toulouse, France, Nov. 2017. Mathematical Methods of Modern Statistics, CIRM Luminy, France, July 2017. Working group on model based clustering group, Paris, France, July 2016. Swiss Statistics day, Bern, Switzerland, April 2015. French annual Chemometrics meeting, Brest, France, Sept. 2013. StatLearn, Bordeaux, France, April 2013. Workshop on ThRee-way methods In Chemistry And Psychology (and other areas) TRICAP, Bruges, Belgium, June 2012. Workshop on Statistical interpretation of Genotype-Environment data (plant breeding), Inra, Jouy en Josas, France, April 2012. French Statistical Society (SFdS), Bruxelle, Belgique, May 2012. (Talk for the Marie-Jeanne Laurent-Duhamel award for my Ph.D. dissertation).

Invited in sessions and Contributed talks

- IMS 2022, Talk in session on Modern Approaches to Missing Data. Online, June 2022, London, UK

- useR!2017 (The R user Conference), Bruxelles, Belgium, July, 2017. French Statistical Society (SFdS), Avignon, France, June, 2017. French Statistical Society (SFdS), Rennes, France, June, 2014. Agro-Stat, Rabat, Maroco, March, 2014. Rencontres R, Lyon, France, June, 2013. Computing & Statistics (ERCIM), London, UK, December, 2013 (Invited in session). Computing & Statistics (ERCIM), Oviedo, Spain, December, 2012. Computational Statistics (COMPSTAT), Cyprus, Aug., 2012. Rencontres R, Bordeaux, France, July, 2012. useR! 2011 (The R user Conference), Warwick, England, August, 2011. German Classification Society (GfKl) and International Federation of Classification Societies (IFCS), Frankfurt, Germany, August, 2011 (Invited in session). International Classification Conference, St Andrew, Scotland, July, 2011. (Invited in session). French Statistical Society (SFdS), Marseille, France, May, 2010. French Statistical Society (SFdS), Bordeaux, France, May, 2009. Agro-stat, Louvain-la-Neuve, Belgique, January, 2009. French Statistical Society (SFdS) and Statistical Society of Canada (SSC), Ottawa, Canada, May, 2008. useR! 2007 (The R user Conference), Iowa State University, Ames, Iowa, August, 2007.

Seminars

- Seminar, Labex NUMEV, Montpellier, February 2023.

- Online Causal Inference Seminar, November, 2022

- Statistics Seminar, Ensae (grande école d'ingénieur in statistics), November 2021

- Statistics Seminar, Montpellier University, November 2021

- Biostatistic and Epidémiology Seminar, Gustave Roussy, Paris, November 2020

- Statistics Seminar, Edinbourgh University, November 2020

- Science des données, inter-établissement Nantes, November 2020

- DataIA, Inria, October 2020

- Van der Schaar lab Seminar, Cambridge, October 2020

- Department of Industrial Engineering, Sharif University of Technology, October 2020

- ESSEC (business school) working group on risk, Paris, June 2020

- Statistics Seminar, ETH Zurich , February 2020

- Statistics Seminar, Cambridge UK, May 2019

- Seminar, Ponts et Chaussés (grande école d'ingénieurs), May 2019

- AI Seminar, Cornell University (Operations Research), March 2019

- Rencontres statistiques lyonnaises, February, 2019

- Statistics seminar, Nice, January, 2019

- Statistics seminar, Bordeaux University, October 2018

- Statistics seminar, MaIAGE, Inrae, Jouy en Josas, March 2018

- Statistics seminar, Washington University, Seatle, February 2018

- Statistics seminar, Paris, Institut Henri Poincaré, January 2018

- Working group in Biostatistics, Department of Biostatistics and Epidemiology University of Massachusetts, Amherst, (remote) 20 July 2017. Seminar, CosmoStat, CosmoClub CEA saclay, 6 April 2017. Statistics seminar, Telecom, France, 23 March 2017. Statistics seminar, University of Economics and Business, Vienna, Austria, 17 March 2017. Statistics seminar, AgroParisTech, France, 13 June 2016. Statistics for Systems Biology seminar, Paris, France, 12 April 2016. Orsay, Statistics seminar, Paris Sud, France, 18 February 2016. INRIA Modal team, Statistics seminar, Lille, France, 14 April 2015. Toulouse School of Economics, Statistic and Econometrics seminar, France, 24 March 2015. Erasmus School of Economics, Econometrics seminar, Rotterdam, Hollande, 12 February 2015. Stanford University, Workshop in Biostatistics, Palo Alto, California, US, 15 January 2015. Malgorzata Bogdan group meeting, University of Wroclaw, Poland, 31 October 2014. Paris Descartes Map5, Statistics seminar, France, 24 October 2014. Montpellier 2 University, Statistics seminar, France, 20 October 2014. Angers University, Statistics seminar, France, 22 September 2014. Institut de mathématiques Bordeaux (IMB), Statistics seminar, France, 6 March 2014. Stanford University, Statistics seminar, Palo Alto, California, US, 23 July 2013. Susan Holmes group's meeting, Stanford University, Palo Alto, California, US, 5 June 2013. Stanford University, Workshop in Biostatistics, Palo Alto, California, US, 23 May 2013. CNAM, Statistics seminar, Paris, France, 16 January 2013. Seminar for researchers of Agrocampus and INRA specialized in milk production and crop science, Agrocampus Rennes, France, 3rd October 2011. Seminar for the psychometric's team of Leiden University, The Netherlands, 20 June 2011. Statistics seminar, IUT de Caen Département STID, France, 22 March 2011. Statistics seminar, équipe Mathématiques et Informatique Appliquées (MIA), INRA, Jouy-en-Josas, France, 8 March 2011. Seminar in statistics for social sciences, Université Paris 1, France, 25 February 2011. Statistics seminar, institut de math'ematiques Bordeaux (IMB), France, 8 March 2010. Statistics seminar, institut de recherche mathématique avancée (IRMA), Strasbourg, France, 7 October 2008. Statistics seminar, institut Elie Cartan, Nancy, France, 13rd June 2008.

Seminars for private companies

- Criteo Tech talk, Paris, France, January 2021

- Google Brain, Tech talk, Paris, France, April 2019

- Tesla Motors data scientists seminar, Palo Alto, California, USA, August 2015

- Adobe Systems Corporate Headquarters, San Jose, California, USA, January 2015

Tutorial in conferences, schools - short lectures

- (2022). Western Swiss Doctoral School in Statistics and Probability. Winter school, Les Diablerets, 4h30

- (2021). AI for Health Winter School, Paris, 2 days

- (2021). Introduction to Causality with JP Vert and A Dieng, Google Brain IA, Paris, 4h

- (2020). Applied Machine Learning Days, EPFL, Switzerland, 1h30

- (2020). Zurich R courses, Switzerland, 2-days

- (2018). Data Science Summer School, École Polytechnique Paris, France, 6h

- (2018). useR!2018, Brisbane, Australia, 3h

- (2018). Ecole d'Aussois. Modélisation Mathématiques et Biodiversité, France, 4h

- (2017). Missing values for ecologists, Biometry and Evolutionary Biology Lab, Lyon, 6h. (2017). Center of Data Sciences, Saclay, 3h. (2016). useR!2016 (The R user Conference), *Massive and missing data*, Stanford, Palo Alto, 3h. (2015). useR!2015 (The R user Conference), *Handling missing values with principal components methods*, Aalborg, Denmark, 3h. (2014). Stanford University, Palo Alto, 2 sessions, (Master and Phd students in Statistics), *Dealing with missing values*, 3h. (2011). Workshop of the French Statistical Society. Ateliers de la SFdS, *Multi-table methods using the software R*. Paris, 3h, (2010). useR!2010 (The R user Conference), Gaithersburg, Maryland, USA, *Clustering and multiway methods*, 3h. (2008). useR!2008 (The R user Conference), *Exploratory data analysis*, Dortmund, Germany, 3h.

Interdisciplinary consortium

- Traumatrix project: Since 2016, I coordinate the Traumatrix project in partnership with the École Polytechnique (IPP), the CNRS, EHESS (École des Hautes Études en Sciences Sociales), Inria, the traumabase group (30 hospitals) and APHP (les hopitaux de Paris) to improve the management of polytrauma patients by leveraging observational data. The project is funded by ARS (Agence Regional de Santé) and by the skill sponsorship of Capgemini Invent (10 data scientists, project managers, data managers). The project is mature and we have launched prospective studies with a mobile application in the ambulances that collects data and predicts the risk of the patient going into hemorrhagic shock.

- ICUBAM (ICU Bed Availability Monitor) project. During the first waves of Covid crisis, I have coordinated a consortium of computer scientists, statisticians and clinicians to create an application to visualize in real time covid+ bed availability (used by 130 ICU wards in 40 départements (regions) covering more than 2, 000 ICU beds).

Software

I am deeply involved in the open source R statistical software community. I organized the useR conference (500 participants) in 2009 in Rennes and I am regularly in the program committees of useR conferences (e.g., Chair of useR!2016 in Stanford). I am also in the board of the French R committee (organization for coordinating R conferences "Les rencontres R") and I am involved in a task Forwards force on behalf of the R Foundation with the aim of increasing the participation of women and under-represented groups in the R community (I am a founding member in 2015). We have worked for instance on the implementation of code of conducts. We organized diversity committees to award scholarships; newbies session, etc.
In Nov. 2016, I have been elected as a member of the R foundation and of the R Foundation Conference Committee.
I use R for teaching, research, development, consulting, etc. R ensures reproducible research and transfer to different domains. I develop and maintain the following libraries:

- FactoMineR dedicated to principal component methods (PCA, Correspondence Analysis for contingency tables, Multiple Correspondence Analysis for categorical data, Multiple Factor Analaysis for multi-blocks data). Users are from the public sector, private sector or research institution from all around the word (4 500 download by day; more than 5 million in total). Google users group and Youtube videos are also available.

- missMDA to perform principal component methods (PCA, MCA, MFA) with missing values and to impute continuous, categorical and mixed data. Multiple imputation is available. (250 download per day; more than 400,000 in total).

- denoiseR to approximate a low-rank matrix from noisy data (Gaussian and Poisson Noise). Singular values shrinkage methods are implemented.

I have also created R-miss-tastic with some colleagues, a platform for missing data to gather and create resources for users: bibliography, courses, tutorials, implementations, pipelines of analysis in R and Python. We have also provided a task-view (to organize all R packages on a topic) on missing values as well as on causal inference (2022).

Visiting researcher, invited stay

- Simon Institute Berkeley, USA. Invited as a research fellows of the causality semester (2022: stay 2 months over 6)

- Stanford University, USA, Statistics department, Host: Trevor Hastie. Visiting researcher (2018: 3 weeks, 2017: 2 weeks, 2016: 2 months, 2015: 3 months, 2014: 6 months, 2013: 5 months)

- Cornell University, USA, Department of Operations Research and Information Engineering, Host: Madeleine Udell (2019: 1 week)

- Geneve University, Switzerland, Mathematics department, Host: Sylvain Sardy (2017: 2 weeks, 2015: 1 week, 2014: 1 week)

- Washington State University, USA, Statístics department, Host: Peter Hoff (2018: 1 week. Visit January 2015: 1 week)

- Rotterdam Erasmus School of Economics, Econometrics, The Netherlands, Host: Patrick Groenen (2015: 2 weeks)

- Los Alamos National Laboratory, USA, Host: Nicolas Hentgarner (2015: 1 week)

- Wroclaw University, Poland, Mathematics department, Host: Gosia Bogdan (2014: 1 month)

- Leiden University, Host: Peter Kronenberg (2011: 3 months)

Teaching

My teaching activities are diverse from the point of view of the teaching (statistics, machine learning, software, projects, etc.), the level of the students (bachelor, master degrees, PhD) and the training involved (école d'ingénieurs, statistics, agronomy, geography, health, social-sciences etc.).
At École Polytechnique I was the head of the Master Data Sciences for Business (2018 - 2020) in collaboration with HEC (Business School).
Here is a list of my main activities:

École Polytechnique (IPP), Saclay (lecture and labs for engineers students, students in the master of data science (M2), master of data science for business (M1)):

- Causal inference and missing values (*20h/year*), 2019-2022.
  Potential outcome framework, (conditional) average treatment effect estimation, causal forests, etc.; EM algorithms, multiple imputation, etc.

- Data Analysis and Unsupervised Learning (*36h/year*), 2019-2020.
  PCA, hierarchical clustering, partition methods, mixture models, PCA with missing values - matrix completion, etc.

- R for Statistics (*64h/year*), 2016-2018.
  Creation of packages, notebooks, visualisation with ggplot2, interactive graphs, Shiny applications, programming statistical methods with R: ex. EM algorithms

- Machine Learning (*36h/year*), 2017-2019.
  Supervised learning with empirical risk minimization, discriminant analysis, tree based methods, k-nearest neighbors, stochastic gradient descent, etc.

- The art of Regression (*18h*), 2016.
  Gaussian linear regression: likelihood, least squares, ridge regression, lasso, logistic regression, etc.

- DataCamp with companies such as Capgemini, (*20h/year*), 2017-2019.

- Research projects, 2017.

EHESS (School of Advanced Studies in the Social Sciences), Paris:

- Co-creation of the data challenge for social scientists lecture (*18h*), 2021-.
  Kaggle type of competition

Agrocampus Ouest (Grande École d'ingénieur in Agronomy), Rennes (lecture and labs for engineers students and for students in a master degree in applied statistics, M1-M2):

- Experimental design, (*24h*), 2015.
  Fractional factorial design, Response Surface Methods, Optimal design

- Principal component methods ("Analyse des données") (*24h/year*), 2014-2015.
  PCA, Multiple Correspondence Analysis, Correspondence Analysis, Multiple Factor Analysis, Tucker3 models (3-ways methods).

- Statistics (*64h/year*), 2007-2014.
  Point estimate, estimation with intervals, tests, ANOVA, etc.

- Missing values in principal components methods, (*3h/year*), 2012-2014.

Continuing education "Formation Continue" at Polytechnique for many companies Orange, SAP, SFR, etc:
Training on Unsupervised/Supervised Learning, Visualisation, DataCamp, programming with R, etc.

Other programs: Master Université Pierre et Marie Curie Statistics, (_6h_), 2018. Master Big Data Dauphine, Paris, (_6h_), 2017. Master in applied statistics for industry and social sciences, University Rennes 2. (_12h/year_), 2014. Master in fisheries, Agrocampus Ouest, Rennes. (_3h/year_), 2013-2014. Master in statistics, National Economics and Statistics Schools (ENSAI). (_12h_), 2012. Bachelor in biology, Pierre Marie Curie University Paris - Agrocampus Ouest, Roscoff. (_30h_), 2012. Master in food science, AgroParisTech, Massy, Paris. (_3h_), 2011. Master in geography, University Rennes 2. (_18h_), 2010. School of Public Health (EHESP), (_18h_), 2010.

Supervision: 6 months students internship working in companies or public sector (4 students/year 2012-2016)

Other responsibilities: Coordination of the joint Master between Polytechnique and HEC data science for business (2016-2018). Member of the department committee École Polytechnique, Center of applied mathematics department. Member of the teaching committee, Agrocampus (2012-2014). Responsible of Statistical teaching for the engineers (L3), Agrocampus. (2011-2014). Coordinator of several statistical courses.

PhD Program:

- Stanford Statistics department. Principal component methods - Multi tables - Missing values, (_24h_), Stat 300, Summer 2015.

Students - Committee

Postdoc:

- Jeffrey Naf, Feb. 2023 -

- Kimia Nadjahi, Nov. 2021 - Nov. 2022. Joint supervision with Claire Boyer. Current position: postdoc at MIT.

- Judith Abecassis, Sept. 2020 - Sept. 2022. Joint supervision with Bertrand Thirion. Current position: junior researcher Inria Saclay.

- Costanza Tortu, Sept. 2020 - Sept. 2021. (grant from Polytechnique's chaire). Current position: Research fellow, Scuola Superiore Sant'Anna, Italy.

- Marine Le Morvan, June 2019 - Sept. 2021. Joint supervision with Gael Varoquaux. Current position: junior researcher Inria Saclay.

- Fabien Laporte, Sept. 2018 - Sept 2019. Joint supervision with Christophe Biernaki and Gilles Celeux.

- Pavlo Mozharovskyi (grant from Henri Lebesgue Mathematical center) Sept. 2015 - Sept. 2016. Current position: Assistant professor, Telecom Paris.

PhD students:

- Supervision of Charlotte Voinot with Bernard Sebastien (grant Phd thesis Cifre Sanofi), Apr. 2023 -

- Supervision of the medical doctor (MD) Tobias Gauss with Pierre Bouzat (MD), Feb. 2023 -

- Participation to the supervision of Lorenzo Gasparollo, PhD student with Mats Stenrud, Sept. 2022 -

- Supervision of the MD Maxime Fosset with Nicolas Molinari and Boris Jung (MD), May 2022 -

- Supervision of Pan Zhao with Antoine Chambaz, Sep 2021 -

- Supervision of Paul Roussel (grant PhD thesis Cifre with Sanofi) with Bernard Sebastien, Oct 2020 - Oct 2021.

- Supervision of Margaux Zaffran (grant PhD thesis Cifre with EDF) with Aymeric Dieuleveut, Yannig Goude, Olivier Ferran. Sept 2020 - . Winner of the Séphora Berrebi Women in Advanced Mathematics & Computer Science.

- Supervision of Benedicte Colnet (grant Inria) with Gael Varoquaux and Erwan Scornet. Sept 2020 - .

- Supervision of Imke Mayer (grant EHESS) with Jean-Pierre Nadal Sept. 2019 - Sept. 2021. Winner of Google PhD Fellowships award in Machine Learning. Current position: Postdoc Charite Institute Berlin.

- Supervision of Aude Sportisse with Claire Boyer. Sept 2018 - Sept 2021. Current position: Postdoc Inria Masai's team Sophia Antipolis.

- Supervision of Nicolas Prost with E. Scornet & G. Varoquaux Sept 2018 - Sept 2019.

- Supervision of Wei Jiang (grant École Polytechnique). Sept 2017 - Sept 2020.

- Supervision Genevieve Robin (grant École Polytechnique) with Eric Moulines, Oct. 2016 - Oct 2019. Winner of the L'Oréal-UNESCO Pour les Femmes et la Science (Jeunes Talents France 2019). Current position, junior researcher, CNRS.

- Supervision of Vincent Audigier with François Husson, Sept 2013 - Sept 2015. Current position: Assistant professor in CNAM, Paris.

Master students supervision (internship):

- Rémi Khelaff, March 2023- August 2023. Manuel Pichon, April 2019 - Aug. 2019. Teresa Alves De Sousa, April 2019 - Aug. 2019. Antoine Ogier, April 2018 - Sept. 2018. Aravinth Chembu, May 2018. Nicolas Prost, April 2018 - Sept. 2018. Aude Sportisse, April 2018 -Sept. 2018. Wei Jiang, April 2017 - Sept. 2017. Tanu July 2017 - Sept. 2017. C. Laroche and B. Jolivet, June-Aug. 2016. Simon Grah, Avril - Aug. 2016. V. Audigier, March - Aug. 2012. K. A Ibrahim, March - Aug. 2011.

Member of PhD committees:

- Reviewer (rapporteur) of the PhD thesis of Lucas Etourneau under the supervision of Nelle Varoquaux and Thomas Burger. December 2023.
- Comité de suivi individuel (CSI) of Etienne Peyrot under the supervision of François Petit.
- Comité de suivi individuel (CSI) of Louis Vincent, under the supervision of by S. Allasonière, 2022 -.
- Reviewer (rapporteur) of the PhD thesis of Elise Dumas, under the supervision of F. Reyal and C.A. Azencott. 2022.
- CSI of Renan Bernard supervised by C. Friguet and N. Courty, 2022 - 2023.
- PhD defense committee of Arthur Chatton under the supervision of Y. Fourchet. 2021.
- PhD defense committee of Tanguy Dubost (medical doctor, intensive care anesthetist). 2021.
- PhD defense committee of Menyssa Cheriffa under the supervision of S. Chevret, M. Resche-Rigon and R. Pirracchio. 2021.
- PhD defense committee of Patricia Dos Santos under the supervision of by J.P. Laurent. 2021.
- PhD defense committee of Gabriel Frisch under the supervision of J.B. Leger and Y. Granvallet. 2021.
- Reviewer (rapporteur) of the PhD of Lilith Faucheux under the supervision of S. Chevret and V. Soumelis. 2021.
- PhD defense committee of Louis Capitaine under the supervision of R. Genuer and R. Thiebaut. 2020.
- Reviewer (rapporteur) of the DTU PhD-student Niels Bruun Ipsen under the supervision of B. Kjaer Ersboll and L. KaiHansen. 2020.
- Reviewer (rapporteur) of the PhD thesis of Lola Etievant under the supervision of V. Viallon. 2020.
- PhD defense committee of Arnaud Gloaguen under the supervision of A. Tenenhaus and V. Frouin.
- PhD defense committee of Marine Zulian under the supervision of M. Lavielle 2020.
- PhD defense committee of Maximilien Baudry under the supervision of C. Robert. 2020.
- PhD defense committee of Diyan Kalainatha under the supervision of I. Guyon. 2019.
- PhD defense committee of Manon Ansart under the supervision of S. Durrleman. 2019.
- Reviewer (rapporteur) of the PhD thesis of Pascaline Descloux under the supervision of S. Sardy. 2019.
- Reviewer (rapporteur) of the PhD thesis of Beyrem Khalfaoui under the supervision of J.P. Vert. 2019.
- Reviewer (rapporteur) of the phD thesis of Timothée Tabouy under the supervision of J. Chiquet and P. Parbillon. 2019.
- Reviewer (rapporteur) of the PhD thesis of Alyssa Imbert under the supervision of N. Villa-Vialaneix. 2018.
- PhD defense committee of Aurore Archimbaud under the supervision of A. Ruiz Gazen. 2018.
- PhD defense committee of Gautier Marti under the supervision of F. Nielsen. 2017.
- PhD defense committee of Pierre-Alexandre Mattei under the supervision of C. Bouyveron. 2017.
- Reviewer (rapporteur) of the phD thesis of Belchin Kostov under the supervision of M. Becue. Spain, 2015.
- PhD defense committee of Anastasios Bellas under the supervision of C. Bouyveron and M. Cottrell. Paris 1, 2014.

Scientific responsibilities

Head of conference organization

- Organization of the online Workshop on *leveraging observational data with machine learning* with the start-up Owkin and the Prairie institute. June 2021. `https://files.inria.fr/leveraging2021/`

- Organization of a online Workshop in International Conference on Machine Learning 2020 (ICML2020), *Learning with missing values*, with G. Varoquaux, P.A. Mattei and J. Frelsen. July 2020. `https://artemiss-workshop.github.io/`

- President of the 1st conference on *missing values and matrix completion MissData*, Rennes, France, June 2015. `http://missdata2015.agrocampus-ouest.fr/`

- President of the international CARME (Correspondence Analysis conference), Agrocampus, Rennes, February 2011. (150 people). `http://carme2011.agrocampus-ouest.fr/`

- Vice-president of the international R conference useR!2009 conference, Agrocampus, Rennes, July 2009 (500 participants). `http://www.r-project.org/conferences/useR-2009/`

Member of conferences scientific committee

- IMS International Conference on Data Science, Lisbonne, Portugal, December 2023.

- Journée de la société Française de Statistique, Bruxelles, Belgium 2023.

- Statlearn, Montpellier, France, April 2023.

- IMS International Conference on Data Science, Florence, Italy, December 2022.

- Mathematical Methods of Modern Statistics 3. Luminy, France, June 2022.

- Mathematical Methods of Modern Statistics 2. Luminy, Marseille, France. June 2020.

- useR!2017, Bruxelles, Belgium, June 2017.

- Les rencontres R, Anglet, France, June 2017.

- Chair of useR!2016, Stanford, CA, USA, July 2016.

- IWSM 2016, International Workshop on Statistical Modelling, Rennes, France, 2016. useR!2015 conference, Aalborg, Denmark, July 2015. MissData, Rennes, France, June 2015. ICCS 2015, Matrix Computational Methods for Modern Data Reykjavik, Iceland, June 2015. French statistical society meeting (SFdS), Lille, June 2015. International SensoMetrics conference, Chicago, United States, July 2014. Second "Rencontres R", Lyon, June 2013, International Sensometrics, Agrocampus, Rennes, July 2012. French Surveys, ENSAI, Rennes, November 2012.

Editorial - reviewing activities

- 2021. Area Chair ICLR

- 2020. Area Chair Neurips

- 2017 - 2020. Associate editor of Journal of Computational and Graphical Statistics

- 2010 - 2017. Associate editor of Journal of Statistical Software

- Invited editor: PNAS, Statistical Science

- Referee for: Journal of the American Statistical Association, Journal of the Royal Statistical Society (C), Journal of Classification, Biometrics, JCGS, JMLR, BMC Bioinformatics, Journal of Statistical Software, Biometrics, Theoretical and Applied Genetics, Environmetrics, Statistical Analysis and Data Mining, Computational Statistic and Data Analysis, Modulad, Advances in Data Analysis and Classification, Statistics and Computing, Npj digital medicine, Statistics in Medicine, R journal, Journal of Statistical Planning and Inference, etc.

- Reviewer for research projects for funding agency. The FWO, Belgium ; the Israel Science Foundation (ISF), "Evaluation of the professorship ad personam", University of Zurich.

Other responsibilities: hiring committees, grants jury, etc.

- Member of the committee to hire two permanent researchers "Chargé de Recherche" (Inrae), Apprentissage statistique pour les sciences du vivant et de l'environnement. 2022.

- Member of the committee to hire an Associate Professor Toulouse 2. 2022.

- Member of the committee to hire the director of the Bio-informatic center of Ecole Nationale Supérieure des Mines de Paris. 2021.

- Member of "Comité du suivi doctoral" Inria Sophia, to allocate PhD grants extensions, etc. 2022.

- Member of the board for Region Occitanie to allocate PhD/postdoc grants in AI for health. 2022.

- Member of the committee to allocate PhD grants for a call Inria-Inserm. 2022.

- Member of the committee to hire an Associate Professor, statistic department, Polytechnique 2017.

- Member of the committee to hire an Associate Professor, Statistics department Dauphine. 2017.

- Member of the committee to hire an Associate Professor, Statistics department, CNAM. 2017. Member of the committee to hire an Associate Professor, Statistics department, Paris 6 University. 2017. Member of the committee to hire an Associate Professor in the Probability and Statistics Department of University of Lille 1. 2015. Member of the committee to hire an Associate Professor in the statistic department of AgroParisTech. 2014.

Other responsibilities: Organisation of seminars, board of societies, etc.

- Organisation of the seminar of the missing data and causality group at Inria (2018 -)

- Co-responsable of the statistics seminar Polytechnique, ENSAE. (2018 - 2020).

- Board Member of the French Statistical Society (Elu au conseil SfdS) in charge of the communication. (2015 - 2018)

- Co-responsable of the Rennes statistics seminar (IRMAR). (2012-2015). Correspondent for the statistical semester 2016 of the Labex Centre de Mathématiques Henri Lebesgue

Outreach, interviews

- Femmes de Tech Â» de lâAcadémie des technologies. 2022.

- Women in Machine Learning profile highlight invitation. 2022.

- Interview for Montpellier University. 2022.

- 2021. Take the twitter account for a week of the the American Statistical Association Committee on Women in Statistics.

- 2020. Interviews for the ICUBAM project (Inria), Polytechnique Insight.

- 2019. Video for the project TrauMatrix. Articles, media cover in in les echos, usine digitale, le cardiologue, pharmaradio

- 2019. Interview in authority magazine Women Leading The AI Industry

- 2018. Table ronde, France is IA conference. 2018: Talk at Women in Machine Learning meetup, Paris. 2018. Blog posts: https://www.r-bloggers.com/multiple-imputation-for-continuous-and-categorical-data/ Blog posts: multivariate studies of the R community 2017. Table ronde, France is IA. 2017, Belgium. Presentation of the diversity initiative in the R community. 2017, Belgium. Animation of the Newbies session at useR! to present the R community to beginners. 2017, Paris. R ladies meetup. 2017, Polytechnique. Participation "Filles et maths : une équation lumineuse". 2014. ENS, Rennes. Participation mathC2+ pour Animath (Association pour l'animation mathématique). Présentation du métier de statisticien pour des lycéens. 2014, INSA, Rennes. Participation "Filles et maths: une équation lumineuse".

Consultancy/contract with companies

- March 2023 - . Consulting for Elixir Health (4h/months, 6 months).

- April 2023 - . Consulting for Quinten Health (2h/months, 6 months).

- Dec. 2020 - Nov. 2022. Member of the Chaire SANOFI-Polytechnique *Numerical Innovation and Data Science for Healthcare*.

- Oct. 2020 - Oct. 2021. PhD thesis Cifre with SANOFI on causal inference.

- Oct. 2020 - PhD thesis Cifre with EDF on price forecast.

- 2019 - 2020. Head of the Chaire Natixis Business analytics and future banking at Polytechnique.

- 2017. Committee for the challenge hackaton *Speed data scientist*, Société Générale - animath. 2014 - 2015. Consultant for Kepler the start-up easyKost. 2014. Consultant for DANONE research, *Assessment of fluid intake across countries around the world: methodological and public health implications*, Paris. 2013 - 2014. Consultant for L'OREAL Paris, *Clustering women of south Africa based on dermatology data, perception data and chemometrics data*. 2012. *Analysis of consumers*, (*3h*), CTIFL (interprofessional center for fruits and vegetables). 2012. Analysis of sensory data, presentation of the software SensoMineR to analyse sensory data. (*12h*), SICAVAC (interprofessional agronomic consulting in viticulture and oenology).