
Curriculum Vitae - January 2021

Julie Josse

Nationality: French.

Date of birth: 10th November 1983.

E-mail: julie.josserennes@gmail.com

Web: <http://juliejosse.com>

Position

Sept 2020 - **Advanced Researcher, INRIA.**

Dec 2020 - **Scientific collaborator + Teaching, CMAP Polytechnique.**

Working Address Inria, Antenne de Montpellier,

860 rue Saint Priest,

34000 Montpellier.

Education - previous positions

- 2016 - August 2020. Professor of Statistics Ecole Polytechnique (IPP) - CMAP. Member of INRIA Team XPOP.
- April 2019 - September 2020. Visiting Researcher, 2 days a week. Google Brain IA, Paris.
- 2016: **Habilitation à diriger des recherches defended 30 August.** Orsay University. “*Contribution to missing values and principal components methods*”. **Committee:** reviewers: G. Celeux (INRIA), J. Friedman (Univ. Stanford), P. Hoff (Univ. Washington); members: A. Ruiz-Gazen (Univ. Toulouse 1), P. Massart (Orsay), E. Moulines (Polytechnique), F. Murtagh (Univ. London), D. Paindaveine (Univ. Bruxelles).
- 2016. Research associate (délégation) team SELECT, INRIA University Paris-Sud, Orsay
- 2011–2015 Associate professor - Statistics department, Agrocampus Ouest (grande école, Ministry of agriculture), Rennes (Brittany, France). (IRMAR - UMR 6625 - CNRS).
- 2007–2010: **PhD thesis in Statistics defended 22 October 2010.** Agrocampus. “*Gestion des données manquantes en analyse exploratoire des données*”. Advisors: J. Pages & F. Husson (Pr, Agrocampus). **Committee:** M. Greenacre (Pr, Barcelona Spain), H.A.L. Kiers (Pr, Groeningen the Netherlands, reviewer), G. Govaert (Pr, Univ. Compiègne France, reviewer), P. Kroonenberg (Pr, Leiden the Netherlands), A. Morin (Ass-pr, IRISA, Rennes France).
Award: best doctoral PhD thesis in applied statistics by the French Statistical Society (Prix Marie-Jeanne Laurent-Duhamel).
- 2006–2010. Engineer - Ingénieur d’études (computer manager) - Agrocampus.
- 2004–2006 Master degree Statistics *summa cum laude* University Rennes II, France.

Publications

Articles

1. Le Morvan, M., J. Josse, E. Scornet. and G. Varoquaux. (2021). What’s a good imputation to predict with missing values?. *Neurips2021, Oral*.
2. Colnet, B., J. Josse, E. Scornet & G. Varoquaux. (2021). Generalizing a causal effect: sensitivity analysis and missing covariates.
3. Mayer, I., J. Josse & the traumabase group (2021). Transporting treatment effects with incomplete attributes. Submitted.
4. P. Sobczyk et al. VARCLUST: clustering variables using dimensionality reduction. Submitted.

5. Moyer, JD, et al. (2020). Trauma reloaded: Trauma registry in the era of data science *Anaesthesia, Critical Care & Pain Medicine*.
6. Le Morvan, M., J. Josse, T. Moreaux, E. Scornet. and G. Varoquaux. (2020). NeuMiss networks: networks: differential programming for supervised learning with missing values. *Neurips2020. Oral*
7. 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.
8. Consortium ICUBAM (2020). ICU Bed Availability Monitoring and analysis in the Grand Est région of France during the COVID-19 epidemic.
9. Debiasing Stochastic Gradient Descent to handle missing values.(2020). Sportisse, A., Josse, J. , Boyer, C. and Dieuleveut, A. *Neurips2020*.
10. Sportisse, A., Boyer, C. and Josse, J. (2020). Estimation and imputation in Probabilistic Principal Component Analysis with Missing Not At Random data. *Neurips2020*.
11. 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*.
12. Mayer, I, Josse, J., Wager, S., Sverdr, E., Moyer, J.D. and Gauss, T. Doubly robust treatment effect estimation with incomplete confounders. 2020. *Annals Of Applied Statistics*.
13. Missing Data Imputation using Optimal Transport. Muzellec, B. Josse, J. , Boyer, C. and Cuturi, M. *International Conference on Machine Learning ICML2020*.
14. Josse, J., Mayer, I, & Vert, J.P. MissDeepCausal: causal inference from incomplete data using deep latent variable models.
15. Descloux, P. Boyer, C., Josse, J. Sportisse, A. and Sardy, S. Robust Lasso-Zero for sparse corruption and model selection with missing covariates. Submitted.
16. Mayer, I, Josse, J., Vialaneix, N., Tierney, N. R-miss-tastic: a unified platform for missing values methods and workflows. 2019. Submitted.
17. M. Bogdan, W. Jiang, J. Josse, B. Miasojedow and V. Rockova. Adaptive Bayesian SLOPE, High dimensional Model Selection with Missing Values. *Journal of Computational and Graphical Statistics*.
18. Josse, J, Prost, N., Scornet, E. & Varoquaux, G. On the consistency of supervised learning with missing values. *In revision in Journal of Machine Learning Research (JMLR)*.
19. Sportisse, A., Boyer, C. and Josse, J. (2019). Imputation and low-rank estimation with Missing Non At Random data. *Statistics & Computing*.
20. G. Robin, O. Klopp, J. Josse, E. Moulines, and R. Tibshirani (2019). Main effects and interactions in mixed and incomplete data frames. *Journal of American Statistical Association*.
21. 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*.
22. G. Robin, Hoi To Wai, J. Josse, O. Klopp and E. Moulines. (2018). Low-rank interactions and sparse additive effects model for large data frames. *NeurIPS*.
23. Josse, J., Husson, F. Robin, G. and Balasubramanian, N (2018). Imputation of mixed data with multilevel SVD. *Journal of Computational and Graphical Statistics*.
24. Mozharovskiy, P., Husson, F. & Josse, J. (2018). Nonparametric imputation by data depth. *Journal of the American Statistical Association*.
25. Hamada et al. (2018). Effect of Fibrinogen administration on early mortality in traumatic haemorrhagic shock: a propensity score analysis. *Journal of Trauma*.
26. 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*.
27. Josse, J. & Reiter, J. (2018). Introduction to the Special Section on Missing Data. *Journal of Statistical Sciences*.

28. Robin, G, Sardy, S., Moulines, E. and Josse, J. (2018). Low rank log-linear models for contingency tables. *Journal of Multivariate Analysis*.
29. Josse, J. & Holmes, S. (2017). Discussion of 50 Years of Data Science. *Journal of Computational and Graphical Statistics*.
30. 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*.
31. Josse, J. , Marin, J.M. & Robert, C.P. Some discussions on the Read Paper "Beyond subjective and objective in statistics" by A. Gelman and C. Hennig.
32. Sobczyk, P., Bogdan, M. & Josse, J. (2017). Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood. *Journal of Computational and Graphical Statistics*.
33. Foulley, JL, Celeux, G and Josse, J. Empirical Bayes approaches to PageRank type algorithms for rating scientific journals. Preprint.
34. Fithian, W. & Josse, J. (2016). Multiple Correspondence Analysis & the Multilogit Bilinear Model. *Journal of Multivariate Analysis*.
35. Husson, F., Josse, J. & Saporta, G. (2016). Jan de Leeuw and the French school of data analysis. *Journal of Statistical Software*.
36. Groenen, P. & Josse, J. (2016). Multinomial multiple correspondence analysis. *On Arxiv: <http://arxiv.org/abs/1603.03174>*.
37. Josse, J., Sardy, S. & Wager, S. (2016). denoiseR a package for low rank-matrix estimation. *On Arxiv: <http://arxiv-web3.library.cornell.edu/abs/1310.6602>*, *Journal of Statistical Software*.
38. 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*.
39. Josse, J. & Wager, S. (2016). Stable Autoencoding: A Flexible Framework for Regularized Low-Rank Matrix Estimation. *Journal of Machine Learning Research*.
40. Audigier, V., Husson, F. & Josse, J. (2016) MIMCA: Multiple imputation for categorical variables with multiple correspondence analysis. *Statistics and Computing*.
41. Josse, J. & Holmes, S (2015). Tests of independence and Beyond. *Statistics Survey*.
42. Audigier, V., Josse, J. & Husson, F. (2015). Multiple imputation for continuous variable using Bayesian PCA. *Journal of Statistical Computation and Simulation*.
43. Josse, J. & Sardy, S. (2015). Adaptive Shrinkage of singular values. *Statistics and Computing*.
44. Josse, J. & Husson, F. (2015). missMDA a package to handle missing values in and with multivariate data analysis methods. *Journal of Statistical Software*.
45. Audigier, V., Husson, F. & Josse, J. (2014). A principal components method to impute mixed data. *Advances in Data Analysis and Classification*.
46. Josse, J., Wager, S. & Husson, F. (2014). Confidence areas for fixed-effects PCA. *Journal of Computational and Graphical Statistics*.
47. Dray, S & Josse, J. (2014). Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology*. **216** (5), 657-667.
48. Josse, J., van Eeuwijk, F., Piepho, 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*. **19** (2), 240-257
49. Verbanck, M. & Josse, J. & Husson, F. (2015). Regularized PCA to denoise and visualise data. *Statistics and Computing*. **25** (2), 471-486.
50. Josse, J. & Timmerman, M.E. & Kiers, H.A.L. (2013). Missing values in multi-level simultaneous component analysis. *Chemometrics and Intelligent Laboratory Systems*. **129**, 21-32.

51. Josse, J., & Husson, F. (2013). Handling missing values in Multiple Factor Analysis. *Food Quality and Preferences*. **30** (2), 77-85.
52. Josse, J., & Husson, F. (2013). Handling missing values in multivariate exploratory data analysis. *Journal of the French Statistical Society (SFdS)*. **153** (2), 79-99. (Paper written for the best Ph.D doctoral thesis prize delivered by the French Statistical Society).
53. Josse, J., Chavent, M., Liquet, B. & Husson, F. (2012). Handling missing values with Regularized Iterative Multiple Correspondence Analysis. *Journal of Classification*. **29** (1), 91-116.
54. Denis, J.B. & Josse, J. (2012). Inferring biadditive models within the Bayesian paradigm. Technical report: R2012. <http://prodinra.inra.fr/record/50125>.
55. Josse, J. & Husson, F. (2011). Selecting the number of components in PCA using cross-validation approximations. *Computational Statistics and Data Analysis*. **56** (6), 1869-1879.
56. Josse, J., Pagès, J. & Husson, F. (2011). Multiple Imputation in PCA. *Advances in Data Analysis and Classification*. **5** (3), 231-246.
57. Husson, F., Josse, J. & Pagès, J. (2010). Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? *Technical report*.
58. Josse, J., Pagès, J. & Husson, F. (2009). Données manquantes en analyse en composantes principales. *Journal de la Soci'et'e Fran'caise de Statistique (SFdS)*. **150** (2), 28-51.
59. Josse, J., Pagès, J. & Husson, F. (2008). Testing the significance of the RV coefficient. *Computational Statistics and Data Analysis*. **53**, 82-91.
60. Lé, S., Josse, J. & Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*. **25** (1), 1-18.

Books

- Husson et al. R pour les statistiques et la data-science (2018), Presses Universitaires de Rennes. <https://r-stat-sc-donnees.github.io/>
- PA. Cornillon, A. Guyader, F. Husson, N. Jégou, J. Josse, M. Kloareg, E. Matzner-Løber, L. Rouvière. (2012). R for Statistics, Chapman & Hall. 320 pages. 800 books sold. <http://www.crcpress.com/product/isbn/9781439881453>.
- PA. Cornillon et al. (2008, 2010, 2012). Statistiques avec R, Presses Universitaires de Rennes. 300 pages. More than 8000 books sold.

Chapter in books

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

Awards and Fellowships

- PRCI (ANR/DFG) ASCAI, PI: Nicolas Verzelen, & Alexandra Carpentier (275Ke).
- Chaire Muse (ANR) (2021). Premedical. Precision medicine by causal learning (300Ke).
- Project SCAI (Sorbonne Center for AI) Apprentissage Statistique pour les hopitaux with Claire Boyer, PI and Boris Muzellec. (70Ke).
- Project 'soutien à la recherche' (Univserity of Montpellier): Real Life Reinforcement Learning with Oldaric-A. Maillard, PI and Benoite de Saporta. (8Ke)
- Independent Research Fund Denmark Grant (2019). Missing data imputation using deep generative models. Project leader Jes Frellsen with P-A Mattei. (400Ke, a PhD student Hugo Senetaire, 1 workshop, mutual visits).
- Capgemini Invent support (2019 -), data for good program, TrauMatrix Project. 10 persons allocated (data/project managers, data-scientists) to create the decision tool to improve the care of trauma patients. (with Pierron-Perles, Charlotte)

- DataIA grant (2018-2020) "Missing data in the big data era" with Gael Varoquaux, INRIA Parietal. 250Ke.
- R consortium grant (2018-2019) "A unified platform for missing values methods and workflows". 10Ke.
- PEPS grant (2017-2019) <http://www.agence-maths-entreprises.fr/a/?q=fr/peps/> for the project handling missing data with heterogeneous data - modeling severe polytrauma patients (with Paris hospital). Money for one internship of 6 months and 1 postdoc.
- AgreenSkills grant (Marie-Curie FP7 COFUND People Program) for mobility research (2013-2014: 12 months). It is an international program co-funded by the European Union designed to increase the research potential of promising researchers. <http://www.agreenskills.eu>.
- Ph.D. dissertation award (Marie-Jeanne Laurent-Duhamel - French Statistical Society), it rewards the quality of a young, french statistician's doctoral thesis. In 2012, it distinguishes a notable contribution to applied statistical research. <http://www.sfds.asso.fr/>

Talks in conferences, workshops, seminars

Conferences keynotes

- AutoML conference. July 2022. Baltimore, USA.
- DAG Stat. Section Compstat. German Consortium Conference in Statistics, March, 2022. Hambourg, Germany.
- Talk PharmScience, Research Development, February, 2022. San Francisco, USA.
- Missing values in Signal Processing, December 2021.
- European Statistical Forum Application of Causal Inference in Drug Development, November 2021.
- Open Data Science Conference, West Virtual, November, 2021. <https://odsc.com/boston/>
- Optimization, Statistics, and Theoretical Machine Learning, Princeton, USA, September, 2020. Online.
- Workshop in the international society for clinical biostatistic, Krakow, Poland, August 2020. Online.
- Workshop Mathematical Methods of Modern Statistics, CIRM, Marseilles, France, June 2020. Online.
- International Federation of Classification Societies, Thessaloniki, Greece, August 2019.
- useR!2019, Toulouse, France, July 2019.
- R in Official Statistics, Bucharest, Romania, May 2019.
- French Statistical Society, Nancy, France, June 2019.
- Statlearn 2019, Grenoble, France.
- R Québec Montreal, Canada, July 2018.
- Insurance, Actuarial Science, Data and Models, Paris, June 2018.

- Older conferences and keynotes. Finnish Workshop on Invariant Coordinate Selection. Toulouse, France, November 2017. / Mathematical Methods of Modern Statistics, CIRM Luminy, France, July 2017. / useR!2017 (The R user Conference), Bruxelles, Belgium, July, 2017. / French Statistical Society (SFdS), Avignon, France, June, 2017. / Working group on model based clustering group, Paris, France, July 2016. **Invited: 1 hour.** / Swiss Statistics day, Bern, Switzerland, April, 2015. **Invited: 1 hour.** / French Statistical Society (SFdS), Rennes, France, June, 2014. / AgroStat, Rabat, Maroco, March, 2014. / French annual Chemometrics meeting, Brest, France, September, 2013. **Invited: 1 hour.** / Computing & Statistics (ERCIM), London, UK, December, 2013. Invited in session/ Rencontres R, Lyon, France, June, 2013./ StatLearn Bordeaux, France, April, 2013. **Invited: 1 hour.** French Statistical Society (SFdS), Bruxelles, Belgique, May 2012. **Invited: 1 hour.** Talk for the Marie-Jeanne Laurent-Duhamel award for my Ph.D. dissertation/ Computing & Statistics (ERCIM), Oviedo, Spain, December, 2012/ Computational Statistics (COMPSTAT), Cyprus, August, 2012/ Rencontres R, Bordeaux, France, July, 2012/ German Classification Society (GfKl) and International Federation of Classification Societies (IFCS), Frankfurt, Germany, August, 2011. **Invited in session**/ useR! 2011 (The R user Conference), Warwick, England, August, 2011/ International Classification Conference, St Andrew, Scotland, July, 2011. **Invited in session.**/ Correspondence Analysis and Related MEthods (CARME), Rennes, France, February, 2011/ French Statistical Society (SFdS), Marseille, France, May, 2010/ French Statistical Society (SFdS), Bordeaux, France, May, 2009/ Agrostat, 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

- Johns Hopkins Mathematical Institute for Data Science seminar, 2022.
- Statistics Seminar, Ensaie, November 2021.
- Statistics Seminar, Montpellier University, November 2021.
- Industry Criteo Tech talk, January 2021.
- Biostatistique et d'Epidémiologie Seminar, Gustave Roussy, November 2020.
- Statistics Seminar, Edinburgh University, November 2020.
- Science des données, inter-établissement Nantes, November 2020.
- DataIA, INRIA, October 2020.
- Van der Schaar lab, Cambridge. October 2020.
- Department of Industrial Engineering, Sharif University of Technology, October 2020.
- ESSEC working Group on risk, Paris, June 2020.
- Statistics Seminar, Zurich, February 2020.
- Statistics Seminar, Cambridge, May 2019.
- Seminar, Ponts et Chaussées, May 2019.
- Google Brain, Tech talk, April 2019.
- AI Seminar, Cornell University (Operations Research), March 2019.
- Kampala R User group (remote), 2019.
- Rencontres statistiques lyonnaises, February, 2019.
- Statistics seminar, Nice, January, 2019.
- Statistics seminar, Bordeaux, October 2018.
- Statistics seminar, MaIAGE, Jouy en Josas, 26 March 2018.
- Statistics seminar, Washington University, Seattle, 28 February 2018.
- Statistics seminar, Paris, IHP, 15 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. / **Industry**. Tesla Motors data scientists seminar, Palo Alto, California, US, 14 August 2015. / 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./ **Industry**. Adobe Systems Corporate Headquarters, San Jose, California, 20 January 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. Workshop on ThRee-way methods In Chemistry And Psychology (and other areas) TRICAP, Bruges, Belgium, June 2-7th 2012/ Workshop on Statistical interpretation of Genotype-Environment data (plant breeding), INRA (French National Institute for Agricultural Research), Jouy en Josas, France, 4 April 2012/ 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 D’epartement STID, France, 22 March 2011/ Statistics seminar, ’equipe 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 mathématiques 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.

Tutorial in conferences - small lectures

- (2022). Western Swiss Doctoral School in Statistics and Probability. (4h30).
- (2021-2022). AI for Health Winter School, Paris (2 days).
- (2020). Applied Machine Learning Days, EPFL (1h30)
- (2020). Zurich R courses (2-days).
- (2018). Data Science Summer School, Polytechnique (6h).
- (2018). useR!2018. Brisbane. (3h).
- (2018). Ecole d’Aussois. Modélisation Mathématiques et Biodiversité. (4h).
- (2017). Missing values for ecologists, Biometry and Evolutionary Biology Lab, Lyon, (6h). / (2017). Missing values. Center of Data Sciences, Saclay, 2 Mars 2017 (3h)/ (2016). Massive and missing data (3h). useR!2016 (The R user Conference), Stanford, Palo Alto, June 27-30th. / Josse, J., & Husson, F. (2015). Handling missing values with principal components methods (3h). useR!2015 (The R user Conference), Aalborg, Denmark, July 1-3th. Josse, J., & Husson, F. (2014). Dealing with missing values (3h). Stanford University, Palo Alto, 2 sessions, 9 and 11 July. (Master and Phd students in Statistics). (2011). Multi-table methods using the software R (3h). Workshop of the French Statistical Society. Ateliers de la SFdS. Paris, 1st April/ Josse, J. & Husson, F. (2010). Clustering and multiway methods (3h). useR!2010 (The R user Conference), Gaithersburg, Maryland, USA, July 20-23 th / Josse, J., Husson, F. & Le, S. (2008). Exploratory data analysis (3h). useR!2008 (The R user Conference), Dortmund, Germany, August.

Project: decision tool for trauma patients

- Traumatrix <https://www.traumatrix.fr/>

Softwares

I am deeply involved in the free R statistical software community. We organized the useR conference (500 participants) in 2009 in Rennes and I am regularly in PC of useR conferences. I use R for teaching, research, development, consulting, etc. It allows us to make the research reproducible and available for the users of different fields. I am working on the packages:

- FactoMineR dedicated to principal component methods (PCA, Correspondence Analysis for contingency tables, Multiple Correspondence Analysis for categorical data, MFA for multi-blocks data). Google users group and you-tube videos 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.
- denoiseR to approximate a low-rank matrix from noisy data (Gaussian and Poisson Noise). Singular values shrinkage methods are implemented.

I am in the board of the French R committee (organization for coordinating R conferences "Les rencontres R") and I am involved in a task force on behalf of the R Foundation with the aim of increasing the participation of women and underrepresented groups in the R community (<http://forwards.github.io/>). We work for instance on implementation of code of conduct.

In nov 2016, I have been elected as a member of the R foundation (<https://www.r-project.org/>) and of the R Foundation Conference Committee.

With Nicholas Tierney and Nathalie Vialaneix, we have created a platform for missing data (<https://rmisstastic.netlify.com/>) and realised task-view on missing values.

During the Covid Crisis, I have been involved in the ICUBAM Project.

- <https://icubam.github.io/> Creation of an application to visualize in real time covid+ bed availabilities.

Visiting Researcher

- Simon Institute Berkeley. Research Fellows of the causality semester (March-April 2022).
- Stanford University, Statistics department. Visiting professor (February 2018: 3w, February 2017: 2w, April 2016, July - August 2015, November 2014 - January 2015, June 2014 - August 2014, April 2013 - July 2013)
- Cornell University. Invited stay March 2019: 1 week.
- Geneve University, Mathematics department. January 2017, March 2015, March 2014.
- Washington State University, Statistics department. Invited February 2018: 1 week. Visit January 2015: 1 week.
- Rotterdam Erasmus School of Economics, Econometrics. Invited stay February 2015: 2 weeks.
- Los Alamos National Laboratory. Invited stay February 2015: 1 week.
- Wroclaw University, Mathematics department. Invited stay October 2014.
- Leiden University April - June 2011.

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 (ecole d'ingénieur, statistics, agronomy, geography, health, social-sciences etc.). The slides, exercices, etc.. are on my webpage.

At polytechnique I was the head of the Master Data Sciences for Business (2018 -) in collaboration with HEC.

Here a list of my main activities:

Polytechnique, Saclay:

- Causal inference (20h), 2019-.

- Data Analysis and Unsupervised Learning (*36h*), 2019-2020.
- R for Statistics (*64h/year*), 2016-2018.
- Machine Learning (*36h*), 2017-2019.
- The art of Regression (*18h*), 2016.
- DataCamp, (*20h/year*), 2017-2019.
- Research projects, 2017

EHESS, Paris:

- Data Science pour les SHS (*18h*), 2021-.

Google Brain IA, Paris:

- Introduction to Causality with JP Vert and A Dieng (*4h*), 2021.

Master degree in applied statistics, Agrocampus Ouest, Rennes:

- Experimental design (lecture), (*24h*), 2015.
- Principal component methods PCA, MCA, CA, etc., (*24h/year*), 2014-2015.
- Statistics (estimator, inference, tests, etc.) , (*24h/year*), 2012-2014.
- Missing values in principal components methods, (*3h/year*), 2012-2014.

Other master: Master UPMC Statistics, (*6h*), 2018. Master Big Data Dauphine, (*6h*), 2017. Master in applied statistics for industry and social sciences, University Rennes 2. (*12h/year*), 2006 - 2014. Master in fisheries, Agrocampus Ouest, Rennes. (*3h/year*), 2013-2014. Bachelor in life and environmental sciences, Agrocampus Ouest, Rennes. (*100h/year*), 2006 - 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.

Responsabilités administratives : Coordination du Master X-HEC (Polytechnique, 2016-) - Conseil de département - Membre élu au Conseil des Enseignants (Agrocampus 2012-2014). Responsable des enseignements de statistique des ingénieurs (L3) Agrocampus. (2011-2014).
Coordinator of several statistical courses from the Master.

PhD Program:

- Stanford Statistics department. Principal component methods - Multi tables - Missing values, (*24h*), Stat 300, Summer 2015.

Students - Committee

Postdoc:

- Work with Kimia Nadjahi, November 2021 -
- Work with Judith Abecassis, September 2020 - .
- Work with Costanza Tortu, September 2020 - September 2021.
- Work with Marine le morvan, June 2019 - September 2021.
- Work with Fabien Laporte, Sept. 2018 - Sept 2019.
- Work with Pavlo Mozharovskyi (grant from Henri Lebesgue Mathematical center) Sept. 2015 - Sept. 2016. Current position: Assistant professor in ENSAI, Rennes.

PhD students:

- Supervision of Pan Zhao with Antoine Chambaz, Sep 2021 -
- Supervision of Paul Roussel (CIFRE SANOFI) with Bernard Sebastien Oct 2020 - Oct 2021.

- Supervision of Margaux Zaffran (CIFRE EDF) with Aymeric Dieuleveut, Yannig Goude, Olivier Ferran. Sept 2020 -
- Supervision of Benedicte Colnet with Gael Varoquaux and Erwan Scornet. Sept 2020 - .
- Supervision of Imke Mayer with Jean-Pierre Nadal Sept. 2019 - Sept. 2021.
- Supervision of Aude Sportisse with Claire Boyer. Sept 2018 - Sept 2021.
- Supervision of Nicolas Prost with E. Scornet & G. Varoquaux Sept 2018 - Sept 2019.
- Supervision of the PhD thesis of Wei Jian. Sept 2017 -.Sept 2020.
- Supervision of the PhD thesis of Genevieve Robin 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, postdoc 'Pont et Chaussés'. Winner of the L'Oréal-UNESCO award for young researchers.
- Supervision of the PhD thesis of Vincent Audigier with François Husson, sept 2013 - sept 2015. Current position: Assistant professor in CNAM, Paris.
- Participation to supervise the PhD thesis of Piotr Sobczyk with Malgorzata Bogdan, 2014 - . Participation to the PhD thesis of Marie Verbanck, 2013.

Master students supervision:

- internship of Manuel Pichon (medical doctor - master data-science), April 2019 - August 2019.
- internship of Teresa Alves De Sousa, April 2019 - August 2019.
- internship of Antoine Ogier, April 2018 - Sept 2018.
- internship of Aravinth Chembu, May 2018.
- internship of Nicolas Prost, April 2018 - Sept 2018/ internship of Aude Sportisse, April 2018-Sept 2018/ internship of Wei Jian, April 2017-Sept 2017 / internship of Tanu July 2017-Sept 2017/ internship of C. Laroche and B. Jolivet, June-August 2016. / Supervision of the internship of Simon Grah, Avril-August 2016. / internships. V. Audigier, March-August 2012. K. A Ibrahim, March-August 2011.

Member of committee:

- PhD defense committee of Arthur Chatton under the supervision of Yohan Fourchet. 2021.
- PhD defense committee of Tanguy Dubost (medical doctor, intensive care anesthetist). 2021.
- Member of the committee to hire the director of the Bio-informatic center of Ecole Nationale Supérieure des Mines de Paris. 2021.
- PhD defense committee of Menyssa Cheriffa under the supervision of Sylvie Chevret, Matthieu Resche-Rigon and Romain Pirracchio. 2021.
- PhD defense committee of Patricia Dos Santos supervised by Jean-Paul Laurent. 2021.
- PhD defense committee of Gabriel Frisch under the supervision of Jean-Benoist Leger and Yves Granvallet. 2021.
- Rapporteur PhD thesis of Lilith Fauchoux under the supervision of Sylvie Chevret and Vassili Soumelis. 2021.
- PhD defense committee of Louis Capitaine under the supervision of R. Genuer and R. Thiebaut. 2020.
- Rapporteur of DTU PhD-student Niels Bruun Ipsen under the supervision of Professor Bjarne Kjær Ersbøll and Professor Lars Kai Hansen. 2020.
- Rapporteur 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.
- Rapporteur PhD thesis of Pascaline Descloux under the supervision of S. Sardy. 2019.
- Rapporteur PhD thesis of Beyrem Khalfaoui under the supervision of J. P. Vert. 2019.
- Rapporteur PhD thesis of Timothée Tabouy under the supervision of J. Chiquet & P. Parbillon. 2019.
- Rapporteur 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.
- Member of the committee to hire an associate professor, statistic department, Polytechnique 2017.
- Member of the committee to hire an associate professor, statistic department Dauphine. Poste MCF, 26e section, 2017.
- Select an associate professor, statistic department, CNAM. Poste MCF, 26e section, 2017.
- Select an associate professor, statistic department, Paris 6. Poste MCF, 26e section, 2016.
- Rapporteur PhD defense committee of Belchin Kostov under the supervision of Monique Becue. Spain, 2015. Member of the committee to hire an associate professor in the Probability and Statistics Department of University of Lille 1. Poste MCF, 26e section, 2015. Member of the committee to hire an associate professor in the statistic department of AgroParisTech. Poste MCF, 26e section, 2014. Member of the PhD defense committee of Anastasios Bellas under the supervision of C. Bouyveron and M. Cottrell. Paris 1, 2014. Member of the PhD thesis committee of Anais Chaland (2013). Museum of natural history.

Scientific responsibilities

Head of conference organization

- Organization of a Workshop on leveraging Real Word data <https://files.inria.fr/leveraging2021/> with okwin and prairie. June 2021.
- Organization of a Workshop in International Conference on Machine Learning 2020 (ICML2020). Learning with missing values, with Gael Varoquaux, Pierre-Alexandre Mattei, Jess Frelsen <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 meeting), 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

- Journée de la société Française de Statistique, Belgium 2023.
- IMS International Conference on Data Science, Italy December 2022.
- Area Chair ICLR 2021.
- Area Chair Neurips 2020.
- Mathematical Methods of Modern Statistics 2. June 2020. Luminy, Marseille. <https://conferences.cirm-math.fr/2146.html>
- useR!2017, Bruxel, Belgium, June 2017. <http://user2017.brussels/>.
- Les rencontres R, Anglet, France, June 2017. <http://angletr2017.com/>.
- chair of useR!2016, Stanford, CA, USA, July 2016. <http://user2016.org/>.
- IWSM 2016, International Workshop on Statistical Modelling, Rennes, France, 2016. <http://www.lebesgue.fr/fr/content/sem2016-iwsm2016/> / useR!2015 conference, Aalborg, Denmark, July 2015. <http://user2015.math.aau.dk> / MissData, Rennes, France, June 2015. <http://missdata2015.agrocampus-ouest.fr/> ICCS 2015, Matrix Computational Methods for Modern Data Reykjavik, Iceland, June 1-3. French statistical society meeting (SFdS), Lille, June 2015. <http://jds2015.sfds.asso.fr>. 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

- 2020-2021 - Area Chair ICLR, 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, etc....
- Reviewer for research projects for funding agency. The FWO, Belgium <http://www.fwo.be>; the Israel Science Foundation (ISF), <http://www.isf.org.il/english/>, "Evaluation of the professorship ad personam", University of Zurich.

Other responsibilities

- Co-responsible 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-responsible of the Rennes statistics seminar (IRMAR, UMR CNRS 6625). (2012-2015) / Correspondent for the statistical semester 2016 of the Labex Centre de Mathématique Henri Lebesgue <http://www.lebesgue.fr>

Membership of Society

- French Statistical Society (SFdS)
- Ellis Society <https://ellis.eu/>
- Institute of Mathematical Statistics (IMS)

Diffusion - Media

- 2021. Take the twitter account for a week of the the American Statistical Association Committee on Women in Statistics. <https://github.com/asacowis/rocur>
- 2020. Effect of Hydrochloroquine. NouvelObs.
- 2019. Video for the project TrauMatrix
- 2019. Media cover: les echos, usine digitale, le cardiologue, pharmaradio
- 2019. Interview in authority magazine Women Leading The AI Industry
- Member of the diversity committee for the conference useR!2019. Award 20 sholarships. Organization of the newbies session.
- 2018, Table ronde, France is IA.
- 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. Animation du R ladies meetup./ 2017, Polytechnique. Participation 'e "Filles et maths : une 'equation lumineuse"./ 2015 - Task force from the R foundation to increase women in R software / 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 for the public or private sector/ Supervision with companies

- Member of "conseil d'administration mgen", groupe vyv. December 2021 -

- Member of the Chaire SANOFI-Polytechnique *Numerical Innovation and Data Science for Healthcare*. December 2020-
- Co-supervised a PhD thesis with SANOFI on causal inference. October 2020 -
- Co-supervised a PhD thesis with EDF on predicting price. October 2020 -
- Supervision of a Master Internship with SAFRAN aircraft engine. April 2020. (Reported)
- In charge of the Chaire Natixis "Business analytics and future banking" for Polytechnique. (2019-2020)
- Training for companies with "Formation Continue at Polytechnique" Orange, SAP, SFR, etc: Training on Supervised Learning, DataCamp, programming with R, etc.
- Committee for the challenge hackaton Speed data scientist, Société Générale - animath. 2017. / Consultant for Kepler consulting and the start-up easyKost. December 2014. Consultant for DANONE research, Paris 2014. Assessment of fluid intake across countries around the world: methodological and public health implications. / Consultant for L'OREAL Paris, 2013-2014. Clustering women of south Africa based on dermatology data, perception data and chemometrics data. / Statistics with the R software, (*6 times 8h*), 2010-2014. Inter - Companies Consulting, researchers and students from Agrocampus Ouest, Rennes. Analysis of consumers, (*3h*), 2012. CTIFL (interprofessional center for fruits and vegetables) - Analysis of sensory data, (*12h*), 2012. SICAVAC (interprofessional agronomic consulting in viticulture and oenology) - Presentation of the software SensoMineR to analyse sensory data, (*4h*), 2011. Ph.D students of a public science and technology institution (INRIA), Rennes.