

MARZIA ANGELA CREMONA

Université Laval
Pavillon Palasis-Prince, local 2449
2325, rue de la Terrasse
Québec (Québec) G1V 0A6 (Canada)

Webpage: marziacremona.com
E-mail: marzia.cremona@fsa.ulaval.ca
Phone: +1 418 656 2131, poste 412525

CURRENT POSITION

Assistant Professor, Department of Operations and Decision systems, Université Laval, from Sept. 2019.

OTHER AFFILIATIONS AND MEMBERSHIPS

Adjunct Assistant Professor, Department of Statistics, The Pennsylvania State University, from Aug. 2019.

Associate Researcher, Population health and optimal health practices axis, CHU de Québec research center, from Nov. 2019.

Member, Center for medical genomics, The Pennsylvania State University, from Dec. 2019.

Regular Member, CIRRELT Interuniversity Research Centre on Enterprise, Networks, Logistics and Transportation, from Dec. 2019.

Regular Member, IID Institute Intelligence and Data, from Jun. 2020

Collaborator Member, CRDM_UL Big Data Research Center, Université Laval, from Jan. 2020.

RESEARCH INTERESTS

I develop statistical techniques for the analysis of large, high-dimensional and complex data – in particular functional data. An important aspect of my research is its collaborative and multidisciplinary nature. Indeed, much of my work is at the interface between statistics and computational biology, and my main application area is in the “Omics” sciences.

RESEARCH EXPERIENCE

- **Bruce Lindsay Visiting Assistant Professor**, Department of Statistics, The Pennsylvania State University, Jul. 2017 – Jun. 2019.
- **Postdoctoral scholar**, Department of Statistics, The Pennsylvania State University, Feb 2016 – Jun. 2017.
- **Visiting scholar**, Department of Statistics and Center for Medical Genomics, The Pennsylvania State University, Oct. 2014 – Jan. 2015.
- **Ph.D. student**, MOX, Department of Mathematics, Politecnico di Milano, Nov. 2012 – Jan. 2016.

EDUCATION

Ph.D. (with Distinction) in Mathematical Models and Methods in Engineering, Politecnico di Milano, 2016.

Thesis: Statistical methods for omics data.

Advisor: Prof. Piercesare Secchi; *Co-advisors:* Prof. Laura M. Sangalli, Prof. Simone Vantini.

M.Sc. in Mathematics, Università degli Studi di Milano, 2011.

B.Sc. in Mathematics, Università degli Studi di Milano, 2009.

FUNDING

NSERC Discovery grant, principal investigator (individual grant), 2020–2025 (102,500 CAD).

SSHRC Insight development grant, co-investigator, 2020–2022 (60,550 CAD).

MEI NovaScience (training initiatives in artificial intelligence), team grant, 2021-2023 (284,241 CAD).

Start-up research grant for new assistant professors, Université Laval, 2019–2022 (30,000 CAD).

Promotion of publication in journals rated A and B, Université Laval, 2020 (2,000 CAD).

AWARDS AND FELLOWSHIPS

Research medal award 2020, Faculty of Business Administration, Université Laval, 2020.

Finalist for ISSNAF (Italian Scientists and Scholars of North America) young investigators award, 2018.

Travel support to attend NRC 2018, 20th Meeting of New Researchers in Statistics and Probability, 2018.
Best presentation of an Early Career Investigator at the joint conference CFE-CMStatistics, 2017.
Travel support to attend 2017 UCLA CGSI Computational Genomics Summer Institute, Long course, 2017.
Postdoctoral travel award, The Pennsylvania State University, 2017.
International mobility award, Politecnico di Milano, 2014.
Ph.D. scholarship in the field of “Genomic Computing”, 2012–2015.

PUBLICATIONS IN PEER-REVIEWED JOURNALS (* co-first authors, † co-corresponding authors)

Submitted

1. **Cremona**, Chiaromonte (Submitted) Probabilistic K-mean with local alignment for clustering and motif discovery in functional data. *arXiv* 1808.04773.
2. Boschi, Di Iorio, Testa, **Cremona**[†], Chiaromonte[†] (Submitted) The shapes of an epidemic: using functional data analysis to characterize COVID-19 in Italy. *arXiv* 2008.04700.
3. Severino, **Cremona**, Dadié (Submitted) COVID-19 effects on the Canadian term structure of interest rates. *SSRN* 3762628.

Published

4. Guiblet*, **Cremona***, Harris, Chen, Eckert, Chiaromonte[†], Huang[†], Makova[†] (2021) Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Research* 49(3): 1497–1516. **Press release**.
5. Chen*, **Cremona***, Qi, Mitra, Chiaromonte[†], Makova[†] (2020) Human L1 transposition dynamics unrevealed with functional data analysis. *Molecular Biology and Evolution* 37(12): 3576–3600. **Press release**.
6. Arbeithuber, Hester, **Cremona**, Stoler, Zaidi, Higgins, Anthony, Chiaromonte, Diaz, Makova (2020) Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. *PLoS Biology* 18(7): e3000745. **Press release**.
7. Di Iorio, Chiaromonte, **Cremona** (2020) On the bias of H-scores for comparing biclusters, and how to correct it. *Bioinformatics* 36(9): 2955–2957.
8. Mei, Arbeithuber, **Cremona**, DeGiorgio, Nekrutenko (2019) A high resolution view of adaptive event dynamics in a plasmid. *Genome Biology and Evolution* 11(10): 3022–3034.
9. **Cremona**, Xu, Makova, Reimherr, Chiaromonte, Madrigal (2019) Functional data analysis for computational biology. *Bioinformatics* 35(17): 2311–2313.
10. Guiblet*, **Cremona***, Cechova, Harris, Kejnovska, Kejnovsky, Eckert, Chiaromonte[†], Makova[†] (2018) Long-read sequencing technology indicates genome-wide effects of non-B DNA on polymerization speed and error rate. *Genome Research* 28: 1767-1778. **Press release**.
11. **Cremona***, Pini*, Cumbo, Makova, Chiaromonte[†], Vantini[†] (2018) IWTomics: testing high-resolution sequence-based “Omics” data at multiple locations and scales. *Bioinformatics* 34(13): 2289-2291.
12. Campos-Sánchez*, **Cremona***, Pini, Chiaromonte[†], Makova[†] (2016) Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis. *PLoS Computational Biology* 12(6): e1004956.
13. **Cremona**, Liu, Hu, Bruni, Lewis (2016) Predicting railway wheel wear under uncertainty of wear coefficient, using universal kriging. *Reliability Engineering and System Safety* 154:49-59.
14. **Cremona**, Sangalli, Vantini, Dellino, Pelicci, Secchi, Riva (2015) Peak shape clustering reveals biological insights. *BMC Bioinformatics* 16:349.

CONFERENCE PROCEEDINGS, ABSTRACTS AND BOOK CHAPTERS

15. Torres-Gonzalez, Arbeithuber, Hester, **Cremona**, Stoler, Higgins, Anthony, Chiaromonte, Diaz, Makova (2020) Duplex sequencing uncovers age-related increase in the frequency of de novo indels in mouse mitochondrial DNA. Abstracts from the 53rd European Society of Human Genetics (ESHG) Conference: e-Posters. *European Journal of Human Genetics* 28(S1): 1007-1008.
16. Eckert, Hile, Guiblet, **Cremona**, Stein, Huang, Chiaromonte, Makova (2020) G-quadruplex sequences are barriers to replicative DNA polymerases and hotspots of mutagenesis. Abstracts from the Environmental Mutagenesis and Genomics Society 51st Annual Meeting. *Environmental and Molecular Mutagenesis* 61(S1): 47-47.
17. **Cremona**, Campos-Sánchez, Pini, Vantini, Makova, Chiaromonte (2017) Functional data analysis of “Omics” data: how does the genomic landscape influence integration and fixation of endogenous retroviruses? In book: *Functional Statistics and Related Fields* (editors: Aneiros, Bongiorno, Cao, Vieu). Springer.

18. **Cremona**, Campos-Sánchez, Pini, Vantini, Makova, Chiaromonte (2016) Functional data analysis at the boundary of “Omics”. *Proceedings of IWSM 2016, 31st International Workshop on Statistical Modelling*.
19. Azzimonti, **Cremona**, Ghiglietti, Ieva, Menafoglio, Pini, Zanini (2015) BarCamp: Technology foresight and statistics for the future. In book: *Advances in Complex Data Modeling and Computational Methods in Statistics* (editors: Paganoni, Secchi). Springer.
20. **Cremona**, Pelicci, Riva, Sangalli, Secchi, Vantini (2014) Cluster analysis on shape indices for ChIP-seq data. *Proceedings of SIS 2014, 47th Scientific Meeting of the Italian Statistical Society*.
21. **Cremona**, Riva, Sangalli, Secchi, Vantini (2013) Clustering ChIP-seq data using peak shape. *Proceedings of SCo 2013, 8th Conference on Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction*.

SOFTWARE

- **IWTomics**: Interval-Wise Testing for Omics Data. *Bioconductor R package* and *Galaxy tool*.
- **SIC-ChIP**: Shape Index Clustering for ChIP-seq peaks. Command line R script.

TEACHING AND SUPERVISION EXPERIENCE

Instructor (full responsibility)

Machine Learning in Business Administration (Winter 2021) Audience: business administration MSc and PhD students. Université Laval.

Special Subjects in Quantitative Methods: Statistical Learning Methods and Applications (Winter 2020) Audience: business administration Ph.D. students. Université Laval.

Probability and Statistics for Business (Fall 2019, Fall 2020) Audience: business administration undergraduates. Université Laval.

Applied Regression Analysis (Fall 2017, Fall 2018) Audience: mixed-major undergraduates. The Pennsylvania State University.

Guest lecturer

Applied Statistics (Fall 2016) Audience: non-statistics graduates. The Pennsylvania State University.

Teaching assistant (practice and laboratory class sessions, grading)

Applied Statistics (Spring 2015) M.Sc. in Electrical Engineering, Politecnico di Milano.

Statistics (Fall 2013) B.Sc. in Energy Engineering, Politecnico di Milano.

Probability and Mathematical Statistics (Spring 2013) B.Sc. in Management Engineering, Politecnico di Milano.

Tutor (student assistance)

Probability and Mathematical Statistics (Spring 2015) B.Sc. in Management Engineering, Politecnico di Milano.

Basics of Statistics and Biomedical Signals (Spring 2015) B.Sc. in Biomedical Engineering, Politecnico di Milano.

Statistics (Fall 2013, Spring 2014) B.Sc. in Energy Engineering, Politecnico di Milano.

Mathematical Analysis 2 (Spring 2011) B.Sc. in Mathematics, Università degli Studi di Milano.

Mathematical Analysis IV (Spring 2010) B.Sc. in Mathematics, Università degli Studi di Milano.

Student supervision

Huy Dang (informal co-supervisor), Ph.D. in Statistics, The Pennsylvania State University, in progress.

Pathé Conté (co-supervisor), M.Sc. in Biostatistics, Université Laval, in progress.

Hibat'Allah Talbi (supervisor), M.B.A. in Business Analytics, Université Laval, 2021.

Laila Bentaleb (supervisor), M.B.A. in Business Analytics, Université Laval, 2021.

Ikram Selmi (co-supervisor), B.Sc. in Applied Mathematical Engineering and Modeling, École Nationale Supérieure d'ingénieurs de Tunis, 2020.

Jacopo di Iorio (co-supervisor), Ph.D. visiting period, The Pennsylvania State University, 2019.

Alice Parodi (co-advisor), M.Sc. in Mathematical Engineering, Politecnico di Milano, 2013.

Student thesis committee

Guillaume Manseau, Ph.D. in Business Administration – Finance, Université Laval, in progress.

Tobia Boschi, Ph.D. in Statistics, The Pennsylvania State University, in progress.

ACADEMIC SERVICE AND OUTREACH ACTIVITIES

Manuscript reviewer, Journal of the Royal Statistical Society: Series C; Annals of Applied Statistics; Statistical Modelling; Sankhya B; Statistical Methods & Applications; Statistical Applications in Genetics and Molecular Biology; Genome Biology and Evolution; Plant Phenomics; IEEE/ACM Transactions on Computational Biology and Bioinformatics; International Review of Financial Analysis.

Session organizer and chair, ERCIM 2021, joint conference CFE-CMStatistics, 2021.
Scientific program committee, ERCIM 2020, joint conference CFE-CMStatistics, 2020.
Abstract reviewer, CSSC 2020, Canadian Statistical Student Conference, 2020.
Session organizer and chair, ERCIM 2019, joint conference CFE-CMStatistics, 2019.
Session chair, ENAR 2019 Spring Meeting, 2019.
Session organizer and chair, ERCIM 2018, joint conference CFE-CMStatistics, 2018.
Session chair, NRC 2018, 20th Meeting of New Researchers in Statistics and Probability, 2018.
Session chair, Workshop on Emerging Methods for Sequence Analysis, 2018.
Committee member, Penn State Postdoctoral Society outstanding postdoc mentor award, 2018.
Judge, PJAS Pennsylvania Junior Academy of Science, 2018.
Journal club organizer, 2017 UCLA CGSI Computational Genomics Summer Institute, 2017.
Committee member, Penn State Postdoctoral Society outstanding postdoc award, 2017.
Judge and mentor, ASA DataFest, 2017-2018.
Judge, PSU Graduate Exhibition, 2017.
Session co-organizer, StaTalk on Biostatistics, 2015.
Conference co-organizer, SCo-BarCamp "Technology Foresight and Statistics for the Future", 2013.
Volunteer researcher, European Researchers' nights, 2013-2015.

INVITED PRESENTATIONS

1. Cremona, Boschi, Di Iorio, Testa, Chiaromonte (Apr. 2021) Statistical learning methods for functional data with applications to the study of COVID-19 in Italy. **CIRRELT webinar**, *online seminar*.
2. Cremona, Chiaromonte, Makova (Mar. 2020) Functional data analysis applications to "Omics" sciences. **Women in Data Science Saguenay**, *Chicoutimi*.
3. Cremona, Chiaromonte, Makova (Dec. 2019) Functional data analysis applications to Omics sciences. **ERCIM 2019, joint conference CFE-CMStatistics**, *London*.
4. Cremona, Chiaromonte, Makova (Dec. 2018) Using Interval-Wise Testing to investigate high-resolution "Omics" data at multiple locations and scales. **ERCIM 2018, joint conference CFE-CMStatistics**, *Pisa*.
5. Cremona, Chiaromonte (Oct. 2018) Probabilistic K-mean with local alignment to locally cluster curves and discover functional motifs. **Workshop on Advances in Functional Data Analysis: cluster, location and shape**, *Rennes*.
6. Cremona (Oct. 2018) Probabilistic K-mean with local alignment for functional motif discovery. **ISSNAF Annual Event** (ISSNAF young investigators award finalists' presentations), *Washington*.
7. Cremona, Chiaromonte (Dec. 2017) Probabilistic K-mean with local alignment for functional motif discovery. **ERCIM 2017, joint conference CFE-CMStatistics**, *London*.
8. Cremona (Jul. 2017) Functional Data Analysis testing and linear modeling for high-resolution "Omics" data. **2017 UCLA CGSI Computational Genomics Summer Institute**, *Los Angeles*.

INVITED DEPARTMENTAL SEMINARS

9. Cremona, Boschi, Di Iorio, Testa, Chiaromonte (Dec. 2020) The shapes of an epidemic: using functional data analysis to characterize COVID-19 in Italy. **Université du Québec à Montréal**, *online seminar*.
10. Cremona, Boschi, Di Iorio, Testa, Chiaromonte (Nov. 2020) The shapes of an epidemic: using functional data analysis to characterize COVID-19 in Italy. **R. Clifton Bailey seminar statistics seminars**, **George Mason University**, *online seminar*.
11. Cremona, Chiaromonte (May 2019) Probabilistic K-mean with local alignment to locally cluster curves and discover functional motifs. **University of Augsburg**, *Augsburg*.
12. Cremona (Feb. 2019) Using functional data analysis to exploit high-resolution "Omics" data. **Clemson University**, *Clemson*.
13. Cremona (Feb. 2019) Using functional data analysis to exploit high-resolution "Omics" data. **American University**, *Washington*.
14. Cremona (Feb. 2019) Using functional data analysis to exploit high-resolution "Omics" data. **University of Canterbury**, *Christchurch*.
15. Cremona (Jan. 2019) Using functional data analysis to exploit high-resolution "Omics" data. **University of Otago**, *Dunedin*.
16. Cremona (Jan. 2019) Using functional data analysis to exploit high-resolution "Omics" data. **Wayne State University**, *Detroit*.
17. Cremona (Jan. 2019) Using functional data analysis to exploit high-resolution "Omics" data. **University of Glasgow**, *Glasgow*.

18. Cremona (Jan. 2019) Using functional data analysis to exploit high-resolution “Omics” data. **University of South Florida, Tampa.**
19. Cremona (Nov. 2018) Using functional data analysis to exploit high-resolution “Omics” data. **Miami University, Oxford.**
20. Cremona, Chiaromonte (Mar. 2017) Discovering motifs in “Omics” signals using local clustering of curves. **Sant’Anna School of Advanced Studies, Pisa.**

CONTRIBUTED PRESENTATIONS AND POSTERS

21. Cremona, Chiaromonte (Mar. 2019) Probabilistic K-mean with local alignment for functional motif discovery. **ENAR 2019 Spring Meeting, Philadelphia.**
22. Cremona, Chiaromonte (Jul. 2018) Probabilistic K-mean with local alignment for functional motif discovery. **JSM 2018, Joint Statistical Meetings, Vancouver.**
23. Cremona (Jul. 2018) Functional data analysis applications in “Omics” sciences. **NRC 2018, 20th Meeting of New Researchers in Statistics and Probability, Burnaby.**
24. Cremona, Chiaromonte (Jul. 2018) Probabilistic K-mean with local alignment for functional motif discovery. **DSSV 2018, Data Science, Statistics & Visualization, Wien.**
25. Cremona, Campos-Sánchez, Pini, Vantini, Makova, Chiaromonte (Jun. 2017) Functional data analysis of “Omics” data: how does the genomic landscape influence integration and fixation of endogenous retroviruses? **IWFOS 2017, 4th International Workshop on Functional and Operatorial Statistics, La Coruña.**
26. Campos-Sánchez, Cremona, Pini, Chiaromonte, Makova (May 2016) Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis (poster). **The Biology of Genomes, Cold Spring Harbor.**
27. Cremona, Pelicci, Riva, Sangalli, Secchi, Vantini (Sept. 2014) ChIP-seq peak shape clustering analysis. **EPIGEN-MiChroNetwork Chromatin Seminar “Gene Regulation through Chromatin Structure”, Milano.**
28. Cremona, Pelicci, Riva, Sangalli, Secchi, Vantini (Jun. 2014) Cluster analysis on shape indices for ChIP-seq data. **SIS 2014, 47th Scientific Meeting of the Italian Statistical Society, Cagliari.**
29. Cremona, Riva, Sangalli, Secchi, Vantini (Sept. 2013) Clustering ChIP-seq data using peak shape (poster). **SCo 2013, 8th Conference on Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction, Milano.**
30. Cremona, Sangalli, Secchi, Vantini (Jun. 2013) Clustering of ChIP-seq data through peak shape. **ABS 2013 Applied Bayesian Statistics School “Bayesian Methods for Variable Selection with Applications to High Dimensional Data”, Como.**

PRESENTATIONS AT HOME INSTITUTION

At Université Laval

31. Cremona, Chiaromonte (Jan. 2020) Probabilistic K-mean with local alignment to locally cluster curves and discover functional motifs. **Statistics Department Seminar.**

At The Pennsylvania State University

32. Cremona (Feb. 2019) Using functional data analysis to exploit high-resolution “Omics” data. **Statistics Department Colloquium.**
33. Cremona (Jun. 2018) Discovering functional motifs in “Omics” curves using probabilistic K-mean with local alignment. **Workshop on Emerging Methods for Sequence Analysis.**
34. Guiblet, Cremona, Cechova, Harris (Oct. 2017) Non-B DNA affects polymerase progression and error rates in sequencers and living cells. **Genomics Seminar.**
35. Cremona (Apr. 2017) Exploiting high-resolution “Omics” data with Functional Data Analysis. **Statistics Department Seminar.**
36. Cremona (Feb. 2017) Functional Motif Discovery for “Omics” curves. **Genomics Seminar.**
37. Cremona (Nov. 2016) Discovering motifs in “Omics” signals using local clustering of curves. **Stochastic Modeling and Computational Statistics Seminar.**
38. Cremona (Oct. 2016) Functional Data Analysis for “Omics” (lighting talk). **9th Annual Postdoctoral Research Exhibition.**
39. Campos-Sánchez, Cremona, Pini, Chiaromonte, Makova (May 2016) Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis (poster). **Center for Medical Genomics Retreat.**
40. Cremona (Apr. 2016) A functional data analysis approach to omics data. **Genomics Seminar.**
41. Cremona (Nov. 2014) Peak shape clustering: an application to GATA-1. **Medical Genomics Seminar.**

At Politecnico di Milano

42. Cremona, Parodi (Jul. 2014) Peak shape cluster analysis reveals novel biological insights. *Workshop on Statistics for Omics*.
43. Cremona (Apr. 2014) Peak shape cluster analysis reveals novel biological insights. *IEO-IIT-PoliMi Joint Meeting on Genomic Computing*.

PARTICIPATION TO CONFERENCES AND SUMMER SCHOOLS

44. Data-driven precision medicine and translational research in the era of big data (May 2020) *virtual conference*.
45. Journée québécoise de valorisation des données CRDM-IVADO (Oct. 2019) *Québec*.
46. Rao Prize Conference (May 2019) *University Park*.
47. ENAR 2019 Workshop for Junior Biostatisticians in Health Research (Mar. 2019) *Philadelphia*.
48. Center for Medical Genomics Retreat (May 2018) *Hershey*.
49. 50th Anniversary Conference, Department of Statistics (May 2018) *University Park*.
50. 2017 Bioinformatics and Genomics Retreat (Sept. 2017) *University Park*.
51. 2017 UCLA CGSI Computational Genomics Summer Institute, Long course (Jul. 2017) *Los Angeles*.
52. Rao Prize Conference (May 2017) *University Park*.
53. 2018 Bioinformatics and Genomics Retreat (Sept. 2016) *University Park*.
54. ISNPS Meeting 2015, Biosciences, Medicine and novel Non-Parametric Methods (Jul. 2015) *Graz*.
55. IWFOs 2014, 3rd International Workshop on Functional and Operatorial Statistics (Jun. 2014) *Stresa*.
56. ABS 2013, Applied Bayesian Statistics School "Bayesian Methods for Variable Selection with Applications to High Dimensional Data" (Jun. 2013) *Como*.
57. NETTAB 2012, Integrated Bio-Search (Nov. 2012) *Como*.

MEMBERSHIPS IN PROFESSIONAL SOCIETIES

SSC Statistical Society of Canada.
ASA American Statistical Association.
IMS Institute of Mathematical Statistics.
Bernoulli Society.
CMStatistics (team FDA: Functional Data Analysis).
SMBE Society for Molecular Biology & Evolution.
ISSNAF Italian Scientists and Scholars of North America Foundation.

WORK EXPERIENCE

Analyst in planning and gas portfolio optimization. A2A Trading S.r.l., Milano, Sept. 2011 – Oct. 2012.

LANGUAGES

Italian (mother tongue), English (fluent), French (intermediate).