MARZIA ANGELA CREMONA

The Pennsylvania State University
517 Wartik Laboratory
University Park, PA 16802 (USA)

Webpage: sites.psu.edu/marziacremona
E-mail: mac78@psu.edu
Phone: +1 814 321 5317

CURRENT POSITION

Bruce Lindsay Visiting Assistant Professor
Department of Statistics, The Pennsylvania State University, July 2017 – Present.

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

Postdoctoral scholar, Department of Statistics, The Pennsylvania State University, Feb. 2016 – June 2017.
Mentors: Prof. Francesca Chiaromonte, Prof. Kateryna D. Makova.
Developed statistical methods and computational tools for functional data, applied to "Omics" data relevant to the study of regional variation in mutation processes and features characterizing the genome.

Visiting scholar, Department of Statistics and Center for Medical Genomics, The Pennsylvania State University, Oct. 2014 – Jan. 2015.
Mentors: Prof. Francesca Chiaromonte, Prof. Kateryna D. Makova.
Developed functional data analysis methods to elucidate the effects of genomic landscape features on integration and fixation of endogenous retroviruses.

Ph.D. student, MOX, Department of Mathematics, Politecnico di Milano, Nov. 2012 – Jan. 2016.
Mentor: Prof. Piecersare Secchi.
Developed statistical methods for the study of genomics and epigenomics data, as part of a large project on "Genomic Computing" involving the European Institute of Oncology, the Italian Institute of Technology and the Politecnico di Milano.
Worked on a project on uncertainty analysis of railway wheel profile wear prediction, using spatial statistics techniques.

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.
Thesis: Wavelet frame construction with few generators.
Advisor: Prof. Maura Salvatori.

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

FUNDING AND AWARDS

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 (* indicates co-first authors)

Submitted and in preparation

- Chen*, **Cremona***, Qi, Mitra, Chiaromonte, Makova (In preparation) Dynamic landscape of human L1 transposition revealed with functional data analysis.
- Mei, Arbeithuber, **Cremona**, DeGiorgio, Nekrutenko (Submitted) A high resolution view of adaptive events. *bioRxiv* 429175.
- **Cremona**, Chiaromonte (Submitted) Probabilistic K-mean with local alignment to locally cluster curves and discover functional motifs. *arXiv* 1808.04773.

Published

- **Cremona**, Xu, Makova, Reimherr, Chiaromonte, Madrigal (2019) Functional data analysis for computational biology. *Bioinformatics*, accepted.
- 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.
- **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.
- Campos-Sànchez*, **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.
- **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.
- **Cremona**, Sangalli, Vantini, Dellino, Pelicci, Secchi, Riva (2015) Peak shape clustering reveals biological insights. *BMC Bioinformatics* 16:349.

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.

CONFERENCE PROCEEDINGS AND BOOK CHAPTERS

- **Cremona**, Campos-Sànchez, 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.
- **Cremona**, Campos-Sànchez, Pini, Vantini, Makova, Chiaromonte (2016) Functional data analysis at the boundary of "Omics". *Proceedings of IWSM 2016, 31st International Workshop on Statistical Modelling.*
- 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.
- **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.*
- **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.*

TEACHING EXPERIENCE

Instructor (full responsibility)

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

Alice Parodi, co-advisor for master thesis "Shape analysis of ChIP-seq profiles", Politecnico di Milano, 2013.

ACADEMIC SERVICE AND OUTREACH ACTIVITIES

Manuscript reviewer, IEEE/ACM Transactions on Computational Biology and Bioinformatics; Statistical Applications in Genetics and Molecular Biology.
Session organizer, ERCIM 2019, joint conference CFE-CMStatistics, 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 Biostistics, 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, 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.*
2. 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.
3. Cremona (Oct. 2018) Probabilistic K-mean with local alignment for functional motif discovery. **ISSNAF Annual Event** (ISSNAF young investigators award finalists' presentations), *Washington*.
4. Cremona, Chiaromonte (Dec. 2017) Probabilistic K-mean with local alignment for functional motif discovery. ***ERCIM 2017, joint conference CFE-CMStatistics****, London*.
5. Cremona (July 2017) Functional Data Analysis testing and linear modeling for high-resolution "Omics" data. ***2017 UCLA CGSI Computational Genomics Summer Institute****, Los Angeles.*
6. Cremona (Mar. 2017) Discovering motifs in "Omics" signals using local clustering of curves. ***Seminar at Sant'Anna School of Advanced Studies****, Pisa.*

CONTRIBUTED PRESENTATIONS AND POSTERS

7. Cremona, Chiaromonte (July 2018) Probabilistic K-mean with local alignment for functional motif discovery. ***JSM 2018, Joint Statistical Meetings****, Vancouver.*
8. Cremona (July 2018) Functional data analysis applications in "Omics" sciences. ***NRC 2018, 20th Meeting of New Researchers in Statistics and Probability****, Burnaby.*
9. Cremona, Chiaromonte (July 2018) Probabilistic K-mean with local alignment for functional motif discovery. ***DSSV 2018, Data Science, Statistics & Visualization****, Wien.*
10. Cremona, Campos-Sànchez, Pini, Vantini, Makova, Chiaromonte (June 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.*
11. Campos-Sànchez, 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.*

12. Cremona, Pelicci, Riva, Sangalli, Secchi, Vantini (Sept. 2014) ChIP-seq peak shape clustering analysis. ***EPIGEN-MiChroNetwork Chromatin Seminar "Gene Regulation through Chromatin Structure"****, Milano.*
13. Cremona, Pelicci, Riva, Sangalli, Secchi, Vantini (June 2014) Cluster analysis on shape indices for ChIP-seq data. ***SIS 2014, 47th Scientific Meeting of the Italian Statistical Society****, Cagliari.*
14. 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.*
15. Cremona, Sangalli, Secchi, Vantini (June 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 The Pennsylvania State University

16. Cremona (June 2018) Discovering functional motifs in "Omics" curves using probabilistic K-mean with local alignment. ***Workshop on Emerging Methods for Sequence Analysis****.*
17. Guiblet, Cremona, Cechova, Harris (Oct. 2017) Non-B DNA affects polymerase progression and error rates in sequencers and living cells. ***Genomics Seminar****.*
18. Cremona (Apr. 2017) Exploiting high-resolution "Omics" data with Functional Data Analysis. ***Statistics Department Seminar***.
19. Cremona (Feb. 2017) Functional Motif Discovery for "Omics" curves. ***Genomics Seminar****.*
20. Cremona (Nov. 2016) Discovering motifs in "Omics" signals using local clustering of curves. ***Stochastic Modeling and Computational Statistics Seminar****.*
21. Cremona (Oct. 2016) Functional Data Analysis for "Omics" (lighting talk). ***9th Annual Postdoctoral Research Exhibition****.*
22. Campos-Sànchez, 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****.*
23. Cremona (Apr. 2016) A functional data analysis approach to omics data. ***Genomics Seminar****.*
24. Cremona (Nov. 2014) Peak shape clustering: an application to GATA-1. ***Medical Genomics Seminar****.*

At Politecnico di Milano

25. Cremona, Parodi (July 2014) Peak shape cluster analysis reveals novel biological insights. ***Workshop on Statistics for Omics****.*
26. 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

27. ENAR 2019 Workshop for Junior Biostatisticians in Health Research (March 2019) *Philadelphia*.
28. Center for Medical Genomics Retreat (May 2018) *Hershey*.
29. 50th Anniversary Conference, Department of Statistics (May 2018) *University Park*.
30. 2017 Bioinformatics and Genomics Retreat (Sept. 2017) *University Park*.
31. 2017 UCLA CGSI Computational Genomics Summer Institute, Long course (July 2017) *Los Angeles*.
32. Rao Prize Conference (May 2017) *University Park*.
33. 2018 Bioinformatics and Genomics Retreat (Sept. 2016) *University Park*.
34. ISNPS Meeting 2015, Biosciences, Medicine and novel Non-Parametric Methods (July 2015) *Graz*.
35. IWFOS 2014, 3rd International Workshop on Functional and Operatorial Statistics (June 2014) *Stresa*.
36. ABS 2013, Applied Bayesian Statistics School "Bayesian Methods for Variable Selection with Applications to High Dimensional Data" (June 2013) *Como*.
37. NETTAB 2012, Integrated Bio-Search (Nov. 2012) *Como*.

MEMBERSHIPS IN PROFESSIONAL SOCIETIES

ASA American Statistical Association.
IMS Institute of Mathematical Statistics.
Bernoulli Society.
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; written and oral), French (basic).

PROGRAMMING SKILLS

Good knowledge of R, Matlab, LaTeX, basic knowledge of C, Java, SAS.

REFERENCES

Prof. Piercesare Secchi (Ph.D. advisor)
Department of Mathematics, Politecnico di Milano
Email: piercesare.secchi@polimi.it
Phone: +39 02 2399 4592

Prof. Francesca Chiaromonte (Postdoctoral mentor)
Department of Statistics, The Pennsylvania State University
Email: fxc11@psu.edu
Phone: +1 814 865 7075

Prof. Kateryna Makova (Postdoctoral mentor)
Department of Biology, The Pennsylvania State University
Email: kmakova@bx.psu.edu
Phone: +1 814 863 1619

Prof. David Hunter (Former department head and teaching supervisor)
Department of Statistics, The Pennsylvania State University
Email: dhunter@stat.psu.edu
Phone: +1 814 865 1348