

Andrea Cappozzo

Assistant Professor of Statistical Science

Curriculum Vitae

2024-06-01

- 📍 Department of Economics, Management, and Quantitative Methods, Università degli Studi di Milano, Italy
- 🏠 andreacappozzo.rbind.io
- ✉ andrea.cappozzo@unimi.it
- 🐦 @AndreaCappozzo
- 🌐 AndreaCappozzo

Research Interests

Mixture models, Model-based clustering and classification, Variable selection, Penalized estimation, Robust statistics, Statistical learning, Health Analytics

Work Experience

- 2024/2 – Pres. **Assistant Professor** (Ricercatore SECS-S/01, tipo B), DEMM, Università degli Studi di Milano
- 2021/4 – 2024/2 **Assistant Professor** (Ricercatore SECS-S/01, tipo A), Department of Mathematics, Politecnico di Milano
- 2017/9 – 2021/12 **Volunteer data scientist**, Heartindata, Milan
- 2020/4 – 2021/3 **Postdoctoral researcher**, Department of Statistics & Quantitative Methods, University of Milano-Bicocca
- 2020/1 – 2020/4 **Freelance Data Scientist**, DCG, Milan
- 2015/9 – 2016/9 **Business analyst and planner**, HP Inc, Barcelona

Teaching Experience

- 2022/2 – Pres. **Lecturer**, MSc course in Biostatistics, Politecnico di Milano
- 2021/9 – Pres. **Teaching assistant and Lecturer**, MSc course in Nonparametric statistics, Politecnico di Milano
- 2021/9 – 2021/9 **Lab with R**, Mixture models module, PhD in Economics, Statistics and Data Science, University of Milano-Bicocca
- 2021/5 – 2021/5 **Lab with R**, Modelos de Mixtura con aplicaciones en R, Universidad Nacional de San Agustín de Arequipa
- 2019/3 – 2021/9 **Teaching Assistant**, BSc courses in Statistics and Statistical Methods, University of Milano-Bicocca
- 2019/3 – 2019/9 **Academic Tutor**, BSc courses in Statistical Models and Computer Science, University of Milano-Bicocca
- 2017/9 – 2018/2 **Teaching Assistant**, BSc courses in Statistics, Bocconi University

Education

- 2020/2 **Ph.D. in Statistics and Mathematical Finance** (Doctor Europaeus) University of Milano-Bicocca
- 2017/2 **Statistical Learning online course** (completed with distinction) Stanford University
- 2015/4 **M.Sc. in Statistical Sciences** (with honors) University of Padua
- 2012/7 **B.Sc. in Statistics and Management** (with honors) University of Padua

Visiting Periods

- 2018/3 – 2019/2 **Visiting PhD student**, Insight Centre for Data Analytics, University College Dublin, Ireland
- 2014/1 – 2014/6 **Exchange Semester**, School of Economics and Management, Tilburg University, The Netherlands

Publications

Refereed journals

- [1] L. Benedetti, E. Boniardi, L. Chiani, J. Ghirri, M. Mastropietro, A. Cappozzo, and F. Denti. “Variational inference for semiparametric Bayesian novelty detection in large datasets”. In: *Advances in Data Analysis and Classification* (2023). ISSN: 1862-5347. DOI: 10.1007/s11634-023-00569-z. <https://link.springer.com/10.1007/s11634-023-00569-z>.
- [2] A. Cappozzo, L. A. García-Escudero, F. Greselin, and A. Mayo-Iscar. “Graphical and Computational Tools to Guide Parameter Choice for the Cluster Weighted Robust Model”. In: *Journal of Computational and Graphical Statistics* 32.3 (2023), pp. 1195-1214. ISSN: 1061-8600. DOI: 10.1080/10618600.2022.2154218. <https://www.tandfonline.com/doi/full/10.1080/10618600.2022.2154218>.
- [3] A. Cappozzo, F. Ieva, and G. Fiorito. “A general framework for penalized mixed-effects multitask learning with applications on DNA methylation surrogate biomarkers creation”. In: *The Annals of Applied Statistics* 17.4 (2023), pp. 3257-3282. ISSN: 1932-6157. DOI: 10.1214/23-AOAS1760. <https://projecteuclid.org/journals/annals-of-applied-statistics/volume-17/issue-4/A-general-framework-for-penalized-mixed-effects-multitask-learning-with/10.1214/23-AOAS1760.full>.
- [4] A. Cappozzo, C. McCrory, O. Robinson, A. Freni Sterrantino, C. Sacerdote, V. Krogh, S. Panico, et al. “A blood DNA methylation biomarker for predicting short-term risk of cardiovascular events”. In: *Clinical Epigenetics* 14.1 (2022), p. 121. ISSN: 1868-7075. DOI: 10.1186/s13148-022-01341-4. <https://doi.org/10.1186/s13148-022-01341-4>. <https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-022-01341-4>.
- [5] A. Casa, A. Cappozzo, and M. Fop. “Group-Wise Shrinkage Estimation in Penalized Model-Based Clustering”. In: *Journal of Classification* 39.3 (2022), pp. 648-674. ISSN: 0176-4268. DOI: 10.1007/s00357-022-09421-z. <https://doi.org/10.1007/s00357-022-09421-z>. <https://link.springer.com/10.1007/s00357-022-09421-z>.
- [6] A. Cappozzo, L. Duponchel, F. Greselin, and T. B. Murphy. “Robust variable selection in the framework of classification with label noise and outliers: applications to spectroscopic data in agri-food”. In: *Analytica Chimica Acta* (2021), p. 338245. ISSN: 00032670. DOI: 10.1016/j.aca.2021.338245. <https://doi.org/10.1016/j.aca.2021.338245>.

- [7] A. Cappelletto, L. A. G. García Escudero, F. Greselin, and A. Mayo-Iscar. "Parameter Choice, Stability and Validity for Robust Cluster Weighted Modeling". In: *Stats* 4.3 (2021), pp. 602-615. ISSN: 2571-905X. DOI: 10.3390/stats4030036. <https://www.mdpi.com/2571-905X/4/3/36>.
- [8] A. Cappelletto, F. Greselin, and T. B. Murphy. "Robust variable selection for model-based learning in presence of adulteration". In: *Computational Statistics & Data Analysis* 158 (2021), p. 107186. ISSN: 01679473. DOI: 10.1016/j.csda.2021.107186. <https://linkinghub.elsevier.com/retrieve/pii/S0167947321000207>.
- [9] F. Denti, A. Cappelletto, and F. Greselin. "A two-stage Bayesian semiparametric model for novelty detection with robust prior information". In: *Statistics and Computing* 31.4 (2021), p. 42. ISSN: 0960-3174. DOI: 10.1007/s11222-021-10017-7. <http://arxiv.org/abs/2006.09012> <https://link.springer.com/10.1007/s11222-021-10017-7>.
- [10] A. Cappelletto, F. Greselin, and T. B. Murphy. "A robust approach to model-based classification based on trimming and constraints". In: *Advances in Data Analysis and Classification* 14.2 (2020), pp. 327-354. ISSN: 1862-5347. DOI: 10.1007/s11634-019-00371-w. <http://link.springer.com/10.1007/s11634-019-00371-w>.
- [11] A. Cappelletto, F. Greselin, and T. B. Murphy. "Anomaly and Novelty detection for robust semi-supervised learning". In: *Statistics and Computing* 30.5 (2020), pp. 1545-1571. ISSN: 0960-3174. DOI: 10.1007/s11222-020-09959-1. <http://link.springer.com/10.1007/s11222-020-09959-1>.

Monographs and refereed conference proceedings

- [1] A. Cappelletto, L. A. García-Escudero, F. Greselin, and A. Mayo-Iscar. "Monitoring Tools in Robust CWM for the Analysis of Crime Data". In: *Building Bridges between Soft and Statistical Methodologies for Data Science*. Ed. by L. A. García-Escudero, A. Gordaliza, A. Mayo, M. A. Lubiano Gomez, M. A. Gil, P. Grzegorzewski and O. Hryniewicz. Cham: Springer International Publishing, 2023, pp. 65-72. ISBN: 978-3-031-15509-3.
- [2] A. Cappelletto, C. Masci, F. Ieva, and A. M. Paganoni. "Model-based clustering of right-censored lifetime data with frailties and random covariates". In: *CLADAG 2023 BOOK OF ABSTRACTS AND SHORT PAPERS*. Ed. by C. Rampichini, M. L. Rocca, P. Coretto, G. Giordano and M. L. Parrella. Pearson Education Resources, 2023, pp. 91-94. ISBN: 9788891935632. <https://it.pearson.com/content/dam/region-core/italy/pearson-italy/pdf/Docenti/Universit%C3%A0/CLADAG-2023.pdf>.
- [3] A. Casa, A. Cappelletto, and M. Fop. "Penalized Model-Based Clustering with Group-Dependent Shrinkage Estimation". In: *Building Bridges between Soft and Statistical Methodologies for Data Science*. Ed. by L. A. García-Escudero, A. Gordaliza, A. Mayo, M. A. Lubiano Gomez, M. A. Gil, P. Grzegorzewski and O. Hryniewicz. Cham: Springer International Publishing, 2023, pp. 73-78. ISBN: 978-3-031-15509-3.
- [4] A. Rossi, A. Cappelletto, and F. Ieva. "Does the Inflation Factor need tuning? Simulation-based adjustment for Outlier Detection via the Functional Boxplot". In: *Book of Short papers SIS 2023*. Milano: Pearson, 2023, pp. 1115-1159. ISBN: 9788891935618AAVV.
- [5] A. Cappelletto, F. Ieva, and G. Fiorito. "Mixed-effects high-dimensional multivariate regression via group-lasso regularization". In: *Book of Short Papers SIS 2022*. Pearson, 2022, pp. 652-657.
- [6] A. Casa, A. Cappelletto, and M. Fop. "Group-wise penalized estimation schemes in model-based clustering". In: *Book of Short Papers SIS 2022*. Pearson, 2022, pp. 534-539.
- [7] A. Casa, A. Cappelletto, and M. Fop. "Group-wise penalized estimation schemes in model-based clustering". In: *Book of Short Papers SIS 2022*. Pearson, 2022, pp. 534-539.
- [8] A. Cappelletto, A. Casa, and M. Fop. "Model-based clustering with sparse matrix mixture models". In: *CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS*. Ed. by G. C. Porzio, C. Rampichini and C. Bocci. Firenze University Press, 2021, pp. 280-283. DOI: 10.36253/978-88-5518-340-6. <https://doi.org/10.36253/978-88-5518-340-6%7D>.
- [9] A. Cappelletto, A. Casa, and M. Fop. "Penalized model-based clustering for three-way data structures". In: *Book of Short Papers SIS 2021*. Pearson, 2021, pp. 758-763.
- [10] A. Cappelletto, L. Duponchel, F. Greselin, and T. B. Murphy. "Robust classification of spectroscopic data in agri-food: first analysis on the stability of results". In: *CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS*. Ed. by G. C. Porzio, C. Rampichini and C. Bocci. Firenze University Press, 2021, pp. 49-52. DOI: 10.36253/978-88-5518-340-6. <https://doi.org/10.36253/978-88-5518-340-6%7D>.
- [11] A. Cappelletto, L. A. G. García Escudero, F. Greselin, and A. Mayo-Iscar. "Exploring solutions via monitoring for cluster weighted robust models". In: *CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS*. Ed. by G. C. Porzio, C. Rampichini and C. Bocci. Firenze University Press, 2021, pp. 284-287. DOI: 10.36253/978-88-5518-340-6. <https://doi.org/10.36253/978-88-5518-340-6%7D>.
- [12] A. Cappelletto and F. Greselin. "Monitoring tools for robust estimation of Cluster Weighted Models". In: *Book of Short Papers SIS 2021*. Pearson, 2021, pp. 1245-1250.
- [13] A. Cappelletto, F. Greselin, and T. B. Murphy. "Robust Model-Based Learning to Discover New Wheat Varieties and Discriminate Adulterated Kernels in X-Ray Images". In: *Statistical Learning and Modeling in Data Analysis*. Ed. by S. Balzano, G. C. Porzio, R. Salvatore, D. Vistocco and M. Vichi. Cham: Springer International Publishing, 2021, pp. 29-36. ISBN: 978-3-030-69944-4. DOI: 10.1007/978-3-030-69944-4_4.
- [14] F. Denti, A. Cappelletto, and F. Greselin. "Outlier and novelty detection for Functional data: a semiparametric Bayesian approach". In: *Book of Short Papers of the 5th international workshop on Models and Learning for Clustering and Classification*. Ledizioni, 2021, pp. 33-38.
- [15] A. Cappelletto, F. Greselin, and T. B. Murphy. "Variable selection for robust model-based learning from contaminated data". In: *Book of Short Papers SIS 2020*. Pearson, 2020, pp. 1117-1122.

- [16] F. Denti, A. Cappozzo, and F. Greselin. "Bayesian nonparametric adaptive classification with robust prior information". In: *Book of Short Papers SIS 2020*. Pearson, 2020, pp. 655-660.
- [17] A. Cappozzo and F. Greselin. "Detecting wine adulterations employing robust mixture of factor analyzers". In: *Statistical Learning of Complex Data*. Springer Berlin Heidelberg, 2019, pp. 13-21. ISBN: 978-3-030-21139-4. DOI: 10.1007/978-3-030-21140-0_2.
- [18] A. Cappozzo, F. Greselin, and G. Manzi. "Predicting and improving smart mobility: a robust model-based approach to the BikeMi BSS". In: *Smart Statistics for Smart Applications 2019 - Book of Short papers*. Milano: Pearson, 2019, pp. 737-742. ISBN: 9788891915108.
- [19] A. Cappozzo, F. Greselin, and T. B. Murphy. "Supervised learning in presence of outliers, label noise and unobserved classes". In: *Book of short papers / Cladag2019*. Centro Editoriale di Ateneo Università di Cassino e del Lazio Meridionale, 2019, pp. 104-107. ISBN: 978-88-8317-108-6.
- [20] A. Cappozzo, F. Ferraccioli, M. Stefanucci, and P. Secchi. "An object oriented approach to multimodal imaging data in neuroscience". In: *Studies in Neural Data Science : Startup Research 2017, Siena, Italy, June 25-27*. Vol. 257. Springer New York LLC, 2018, pp. 57-73. ISBN: 9783030000387. DOI: 10.1007/978-3-030-00039-4_4.
- [21] A. Cappozzo, F. Greselin, and T. B. Murphy. "Robust Updating Classification Rule with applications in Food Authenticity Studies". In: *Book of Short Papers SIS 2018*. Pearson, 2018, pp. 493-499. ISBN: 9788891910233.
- [22] A. Cappozzo, F. Greselin, and T. B. Murphy. "The role of trimming and variable selection in robust model-based classification for food authenticity studies". In: *COMPSTAT 2018 Book of Abstracts*. COMPSTAT and CRoNoS, 2018, pp. 35-35. ISBN: 978-9963-2227-3-5.
- [23] F. Greselin and A. Cappozzo. "Wine authenticity assessed via trimming". In: *Book of short papers / Cladag2017*. Mantova: Universitas Studiorum Srl, 2017. ISBN: 9788899459710.

Presentations

Invited

- 2024/6 **SIS 2024 - The 52nd Scientific Meeting of the Italian Statistical Society**, University of Bari Aldo Moro, Italy
- 2024/5 **I Seminario Internacional de Estadística Aplicada**, Universidad Nacional de San Agustín de Arequipa, Peru
- 2023/12 **CMStatistics 2023**, HTW Berlin, University of Applied Sciences , Germany
- 2023/9 **Scientific meeting CLADAG 2023**, University of Salerno, Italy
- 2022/12 **CMStatistics 2022**, Senate House University of London, United Kingdom
- 2022/9 **10th International Conference on Soft Methods in Probability and Statistics**, University of Valladolid, Spain
- 2022/7 **17th conference of the International Federation of Classification Societies**, University of Porto, Portugal
- 2021/9 **22nd European Young Statisticians Meeting**, virtual conference
- 2019/9 **Scientific meeting CLADAG 2019**, University of Cassino and Southern Lazio, Italy
- 2018/8 **COMPSTAT 2018**, Unirea Hotel, Iasi, Romania

Seminars

- 2024/3 **DAIS group seminar series**, Ca' Foscari University of Venice, Italy
- 2022/2 **Mox seminar series**, Politecnico di Milano, Italy
- 2021/5 **Mox seminar series**, Politecnico di Milano, Italy
- 2020/9 **IR-group meeting**, virtual seminar
- 2019/12 **Mock talk II**, University of Milano-Bicocca, Italy
- 2019/7 **Mock talk I**, University of Milano-Bicocca, Italy
- 2018/10 **Working Group on Statistical Learning**, University College Dublin, Ireland

Contributed

- 2023/6 **SIS 2023 - Statistical Learning, Sustainability and Impact Evaluation**, Università Politecnica delle Marche, Italy
- 2021/9 **Scientific meeting CLADAG 2021**, virtual conference
- 2021/6 **SIS 2021- 50th Meeting of the Italian statistical society**, virtual conference
- 2021/2 **e-CHIMIOMETRIE 2021**, virtual conference
- 2020/9 **MBC² Workshop on Models and Learning for Clustering and Classification**, virtual conference
- 2020/6 **e-Rum 2020**, virtual conference
- 2019/12 **CMStatistics 2019**, Senate House University of London, United Kingdom
- 2019/6 **Smart Statistics for Smart Applications - SIS conference**, Catholic University of the Sacred Heart, Italy
- 2017/9 **Scientific meeting CLADAG 2017**, University of Milano-Bicocca, Italy

Poster/software sessions

- 2023/8 **ISCB2023 – International Society for Clinical Biostatistics**, University of Milano-Bicocca, Italy
 2021/10 **Working Group on Model-Based Clustering**, Athens, Greece
 2021/4 **International workshop on Spectroscopy and Chemometrics**, virtual conference
 2021/1 **DS³ Data science summer school**, virtual conference
 2020/7 **Young-ISA Twitter Poster Conference**, virtual conference
 2019/7 **Working Group on Model-Based Clustering**, Wirtschaftsuniversität Wien, Austria
 2018/9 **MBC² Workshop on Model-Based Clustering and Classification**, University of Catania, Italy

Summer Schools and Workshops

- 2018/4 **Robust Statistics: Foundations and Recent Developments**, University of Milano-Bicocca, Italy
 2017/11 **Introduction to Functional Data Analysis**, Università degli Studi di Bergamo, Italy
 2017/6 **Startup Research**, Certosa di Pontignano, Italy
 2017/5 **International Summer School on Classification and Data Analysis**, University of Bologna, Italy

Awards

- 2023/9 **Young Researcher Paper Award**, CLADAG 2023, University of Salerno, Italy
 2018/9 **Best poster presentation**, MBC² Workshop on Model-Based Clustering and Classification, Catania
 2017/9 **Member of the third winning team**, Young CLADAG - Data science competition, Politecnico di Milano
 2017/6 **Member of one of the four winning teams**, Stats Under the Stars 3 - Data science competition, University of Florence

Theses supervision

Master's Theses

2022/12	Anna Iob	Politecnico di Milano	Text mining, Personalized medicine
2023/05	Annachiara Rossi	Politecnico di Milano	Robust statistics, Functional Data Analysis
2023/10/4	Giulia Montani	Politecnico di Milano	Model-based classification, Noisy labels, Ensemble learning
2023/10/4	Luca Caldera	Politecnico di Milano	Mixed-effects models, Multilevel logistic cluster-weighted model
2023/10/4	Davide Zaltieri, Luca Panzeri	Politecnico di Milano	Robust statistics, Model-based clustering
2023/10/4	Mirko Luigi Carlesso	Politecnico di Milano	Basketball analytics, Tree-based methods, Geostatistics
2023/12/22	Lorenzo Angiolini	Politecnico di Milano	Survival Analysis, Frailty models, Model-based clustering
2024/04/10	Francesca Salvatore	Politecnico di Milano	Multitask learning, Mixed-effects models, Penalized estimation
2024/04/10	Greta Campese	Politecnico di Milano	Mixed-effects models, Multilevel logistic cluster-weighted model
2024/04/10	Elisa Borrini	Politecnico di Milano	Model-based clustering, Penalized estimation, Covariance objects
Ongoing	Alessio Frezza	Politecnico di Milano	Mixed-effects models, Multilevel logistic cluster-weighted model

Service to profession

Referee service

Advances in Data Analysis and Classification, BMC Bioinformatics, Bioinformatics, Computational Statistics, Computational Statistics and Data Analysis, IEEE Transactions on Knowledge and Data Engineering, Journal of Computational and Graphical Statistics, Metron, Scientific reports, Statistical Methods & Applications, Statistical Modelling, Statistics and Computing, Statistics in Medicine, Risk Analysis.

Membership

Italian Statistical Society

Computer skills

R (advanced), latex (advanced), quarto (advanced), bash (intermediate), C++ (intermediate), git (intermediate), python (intermediate), mathematica (basic), matlab (basic), SQL (basic).

Languages

Italian (mother tongue), English (proficient), Spanish (good knowledge)