Alessandra Cabassi

✉ sandy.cabassi@gmail.com
🏠 sandycabassi.com
🐦 @SandyCabassi
 acabassi

Experience

Sep 2020 – now
Data scientist, *Google*, Zurich, Switzerland.
- May 2022 – now: YouTube – Data Science team.
- Sep 2020 – April 2022: Google Ads – Analytics, Insights, and Measurement team.

Jun – Dec 2019
Data science intern, *Google*, London, UK.
Team: Advanced Measurement Technologies. Host: Dr Matthew Pearce. Co-host: Dr Georg M. Goerg.
Implemented large-scale multivariate time series models and anomaly detection tools for ad monitoring, with the aim of reducing computational cost as well as false positive alerts (Python, TensorFlow).

Jun – Sep 2018
Research intern, *The Alan Turing Institute*, London, UK.
Supervisors: Dr Anthony Lee, Dr Ioannis Kosmidis, Dr Rajen Shah, and Dr Yi Yu.
Benchmarked machine learning libraries such as Apache Spark and TensorFlow for high-performance, large-scale linear models on the Institute's supercomputer (Python). Worked in close collaboration with software engineers from Cray EMEA Research Lab.

Jan – May 2018
Teaching assistant, *University of Cambridge*, UK.
Supervised 2nd-year Mathematics students for the Statistics module at four Cambridge Colleges.

Oct – Dec 2015
Visiting student, *Statistical Laboratory, University of Cambridge*, UK.
Supervisors: Dr Davide Pigoli and Prof Piercesare Secchi.
Developed statistical methods for hypothesis testing on covariance operators of functional data, implemented and published the corresponding software, applied the novel methodology to a study in evolutionary biology (R).

May – Sep 2013
Research intern, *INRIA*, French institute for Computer Science and Automation, Nice, France.
Supervisors: Dr Paola Goatin and Prof Nicola Parolini.
Implemented mathematical traffic flow models based on partial differential equations, validated them through extensive numerical simulations, and calibrated them with data provided by industrial partners (Matlab).

Education

2016 – 2020
PhD Biostatistics, *MRC Biostatistics Unit, University of Cambridge*, UK.
Supervisors: Dr Paul DW Kirk & Dr Chris Wallace. Advisor: Prof Sylvia Richardson.
Main research interests: Bayesian Methods, Kernel Methods, Model-Based Clustering, Variational Inference.
- Designed and implemented statistical and machine learning methods for integrative clustering of heterogeneous data, with a focus on datasets with small number of observations and large number of features (R, Matlab).
- Collaborated with clinicians and bioinformaticians to gain insights from novel genomic datasets.
- Selected as one of 28 early-career academics in Europe for cross-disciplinary research project.

2013 – 2016
MSc (Laurea Magistrale) Engineering Mathematics, *Politecnico di Milano*, Italy, 1st.
Main subjects: Statistics, Optimisation, Game Theory, Risk Analysis, Artificial Intelligence.

2011 – 2013
MSc (Diplôme d'Ingénieur) Multidisciplinary Engineering, *École Centrale Nantes*, France, 2:1.
Main subjects: Elements of Civil, Electronic, and Energy Engineering, Project Management, Communication.

2009 – 2011
BSc (Laurea Triennale) Engineering Mathematics, *Politecnico di Milano*, Italy.
Main subjects: Advanced mathematics, Numerical analysis, Computer science.

Publications

2022
- Seyres, D., Cabassi, A., … Frontini, M. *Transcriptional, epigenetic and metabolic signatures in cardiometabolic syndrome defined by extreme phenotypes* Clin Epigenet 14, 39 (2022) doi:10.1186/s13148-022-01257-z.

2020
- Cabassi, A., Richardson, S., and Kirk, P.D.W. *Kernel learning approaches for summarising and combining posterior similarity matrices.* arχiv preprint. arxiv:2009.12852
- Cabassi, A., and Kirk, P. D. W. *Multiple kernel learning for integrative consensus clustering of genomic datasets.* Bioinformatics, btaa593. doi:10.1093/bioinformatics/btaa593.
- Cabassi, A., Seyres, D., Frontini, M., and Kirk, P. D. W. *Two-step penalised logistic regression for multi-omic data with an application to cardiometabolic syndrome* arχiv preprint. arxiv:2008.00235

2018 ○ Cabassi, A., Casa, A., ... Farcomeni, A. *Three testing perspectives on connectome data.* In: Canale A., et al. Studies in Neural Data Science. Springer Proceedings in Mathematics & Statistics, vol 257.

2017 ○ Cabassi, A., Pigoli, D., Secchi, P., and Carter, P. A. *Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology.* Electronic Journal of Statistics: Vol. 11, No. 2: 3815-3840, doi:10.1214/17-EJS1347.

2013 ○ Cabassi, A. and Goatin, P. *Validation of traffic flow models on processed GPS data.* Research report RR-8382, INRIA, hal:00876311.

Open source software

klic   Kernel Learning Integrative Clustering, *R package hosted on CRAN*, github.com/acabassi/klic.

coca   Cluster-Of-Clusters Analysis, *R package hosted on CRAN*, github.com/acabassi/coca.

fdcov   Functional Data COVariance operators, *R package hosted on CRAN*, github.com/acabassi/fdcov.

Skills

Languages	In order of proficiency: R, Python, C / C++.	Libraries	TensorFlow, Apache Spark.
Software	Matlab, Microsoft Office.	OS	macOS, Linux, Windows.

Selected academic talks

Invited ○ *Statistical methods for multi-omic data integration.*
2021 World Meeting of the International Society for Bayesian Analysis, held virtually, 06/2021.

○ *Multiple kernel learning for integrative clustering in genomic precision medicine.*
11th International Conference of ERCIM on Computational & Methodological Statistics, Pisa, Italy, 12/2018.

○ *Permutation tests for functional and object data.*
4th conference of the International Society for Nonparametric Statistics, Salerno, Italy, 06/2018.

Contributed ○ *Kernel learning approaches for summarising and combining MCMC samples from the posterior distributions of Bayesian clustering models.*
Statistical Methods for Post Genomic Data workshop, Montpellier, France, 01/2018.

○ *Multiple kernel learning for integrative clustering of 'omics data.*
Biometrika workshop, Cambridge, UK, 11/2017.

Other ○ *High-performance, large-scale regression*, joint talk with J. Wang (Newcastle University), 09/2018.

○ *Three testing perspectives on connectome data*, joint talk with A. Casa (University of Padua), 07/2018.

Scholarships

2017   Funding for high-cost training at University of Washington, *MRC Doctoral Training Partnership*, UK.

2016   3-year scholarship for PhD at University of Cambridge, *MRC Biostatistics Unit*, UK.

2016   Scholarship for Master's thesis (65 available for university of 42k+ students), *Politecnico di Milano*, Italy.

2012   1-year scholarship for outstanding international students, *Pays de la Loire Region*, France.

2011   1-year Erasmus scholarship for double degree, *Top Industrial Managers for Europe Association*.

Languages

Italian	Native speaker.	**English**	Fluent, *C2 level: IELTS 7.5.*
French	Fluent, *C1 level: TFI 950/1000.*	**Spanish**	Fluent, *C1 level: CULP 930/1000.*

Selected student societies

Since 2019   **Mentor** at *Lead The Future*, Nonprofit providing 1:1 mentoring to high-potential Italian students.
Mentoring 2 undergraduate and 3 graduate students at top European universities and institutions.

2014   **President** of *AIM*, Engineering Mathematics Society, Politecnico di Milano.
Lead 9-people committee, restructured the society. It now has 1300+ members and 10+ industrial partnerships.