# Soel Micheletti

*Curriculum vitae*

*Zurich, CH*
*+41 76 404 06 04*
*soel.micheletti@hotmail.it*

## Education

2022-2023 **Harvard University**, *Visiting Graduate Student.*

Master's Thesis on the inference of Gene Regulatory Networks, supervised by John Quackenbush.

2020-2023 **ETH Zurich**, *MSc. Data Science.*

Focus on Machine Learning and Statistics, minor in Finance.

2017-2020 **ETH Zurich**, *BSc. Computer Science.*

With honours, top 15%.

## Professional Experience

2023-present **Software Engineer**, *Oepfelbaum*, Zurich, Switzerland.

Development of financial applications in Python.

- Fast pricing project: developed robust and accurate ML algorithms for fast pricing of barrier structured products.
- Beacon POC project: worked within an international team (Zurich and London) on assessing the Beacon platform as a real-time position keeping and risk management system for a large book of equity options, warrants, and minifutures.
- Mainly working in Python and Pytorch.

2022-present **Research Fellow**, *Harvard University*, Boston, USA.

Remote as a Research Collaborator from October 2023.

- Design of algorithms for the inference and analysis of large complex graphs, with the goal of unveiling mechanisms involved in the development and progression of cancer.
- Open source contributions to netZooPy, netZooR, and netbooks.
- Mainly working in Python, R, and Pytorch.

2022 **Software Engineering Intern**, *Ergon Informatik*, Zurich, Switzerland.

- Refactored a large Java library used for caching of web server requests towards a new caching framework (from Guava to Caffeine). Extensive unit testing and integration testing. Working within an Agile framework.
- Extended the frontend dashboard of the testing platform to support scanning of QR codes.
- Mainly worked in Java, RxJava, Kotlin, and Angular.

2019-2022 **Teacher Assistant**, *ETH Zurich*

See dedicated section.

2016 **Frontend Developer**, *Wonews SA.*, Locarno, Switzerland

- Developed the company's website as a part-time job during high-school.
- Mainly worked in Angular.

## Selected Projects

Summary of projects I completed as part of my studies or after graduation. For a comprehensive overview feel free to check here.

2023 **Research project on batch correction in gene co-expression analysis**

Commonly used batch correction methods account for differences in mean and variance between different groups or batches. However, this is not sufficient for gene co-expression analysis. We drew attention to the problem by showing the existence of residual spurious effects after standard batch correction and introduced COBRA, a method based on matrix factorization to effectively address this gap.

2022-2023 **Thesis on Gene Regulatory Networks Inference**

Developed GIRAFFE, a scalable matrix-factorization based algorithm. By integrating multiple sources of prior knowledge to guide the optimization of a non-convex objective function, GIRAFFE outperforms state-of-the-art methods in inferring interactions between transcription factors and genes.

Spring 2022 **AI Center Project in Machine Learning Research**, *ETH Zurich*

Analysis and open source implementation of a simple algorithm for Bivariate Causal Discovery via Conditional Divergence.

Fall 2021 **Research in Data Science: *Unravelling Stroke Recovery***, *Joint research with the Lucerne Canton Hospital*

Applied state-of-the-art algorithms to estimate undirected graphical models and gain insights on how different recovery categories influence each other; implemented a recurrent neural network to predict recovery patterns.

Spring 2021 **3D Hand Pose Recognition**

Implemented an Encoder-Decoder architecture in Pytorch to efficiently predict heat-maps of the joints.

Fall 2020 **Inference from Medical Data**

- Designed a complete classification pipeline on a MRI dataset (imputation, outlier detection, feature engineering, selection and tuning of the best algorithms to stack in the final model).
- Developed robust algorithms to handle highly imbalanced datasets.

Spring 2020 **Thesis on Evolutionary Algorithms**

- Formally proved results about the performance of evolutionary/ genetic algorithms on relevant benchmark functions.
- MATLAB simulations to interpret the results and propose better performing variants of the algorithm.

2016-2017 **Research on Stochastic Simulation**

- Designed a Monte Carlo algorithm to estimate the area of geographical surfaces and implemented a GUI in Java for end-users.
- Special prize *Odd Fellows* and nominated best project in Computer Science at the 2018 Taiwan International Science Fair.

## Publications

**BioRxiv** **S. Micheletti**; Daniel Schlauch; John Quackenbush; Marouen Ben Guebila, *Higher-order correction of persistent batch effects in correlation networks*, BioRxiv (under review).

**RECOMB** **S. Micheletti**; Alexander Marx; Julia Vogt; John Quackenbush; Jonas Fischer; Panagiotis Mandros, *Gene-level Inference of Regulatory effects As Factorizations of Functions of Expressions (GIRAFFE)*, Research in Computational Molecular Biology.

**RECOMB** **S. Micheletti**; Daniel Schlauch; John Quackenbush; Marouen Ben Guebila, *Higher-order correction of persistent batch effects in correlation networks (short)*, Research in Computational Molecular Biology.

**JAHA** JM. Veerbeek; C. Hutter; **S. Micheletti**; S. Riedi; E. Bianchi; B. Ottiger; N. Maaijwee; T. Vanbellingen; T. Nyffeler, *Profiling Daily Life Performance Recovery in the Early Subacute Phase After Stroke Using a Graphical Modeling Approach*, Journal of the American Heart Association.

## Selection of Awards, Achievement, and Honours

2023 **LeadTheFuture**

Among the few students selected to be mentees for LeadTheFuture, a leading mentorship non-profit organization for students in STEM, with an acceptance rate below 20%.

2021 **Distinguished delegate at the NMUN conference**, *New York*, USA.

Recognized to the top 20% candidates based on team-work attitude, diplomatic capabilities, and quality of the proposed solution.

2018 **First prize in Computer Science**, *Taiwan International Science Fair*, Taipei.

2017 **Valedictorian**, *Liceo Cantonale di Locarno.*

**Awards:** best student of the year, best bilingual matura, Deutsche Bank award.

2017 **Special Prize *Odd Fellows***, *National Competition of Schweizer Jugend Forscht*, Bern, Switzerland.

Recognized to the top STEM projects in the most prestigious scientific fair for high school students in Switzerland.

## Teaching

Parallel to my studies, I was highly involved in teaching. As Albert Einstein said, *"If you can't explain it simply, you don't understand it well enough."* Teaching pushed me to think to fundamental concepts and explain them as clearly as possible, which ultimately made me understand the topics even better. In many cases, teaching assistants pick their favorite course and stick to it over the years. I did it differently, working on multiple areas: from Theoretical Computer Science to Software Engineering, and also a bit of Computer Networks.

Spring 2022 **Computer Networks**, *ETH Zurich*

Graded exercise submissions for about 300 students.

Fall 2021 **Algorithms and Data Structures: intensive course**, *ETH Zurich*

Wrote a script containing the whole scope of the course; held a week of intensive sessions to assist the students in the exam preparation, providing them intuitive explanations of algorithms and tricks to perform well.

Fall 2021 **Algorithms and Data Structures**, *ETH Zurich*

Held weekly exercise sessions for about 20 students; graded their exercise sheets; answered the students' questions in the online chat during lectures.

Spring 2021 **Computer Networks**, *ETH Zurich*

Graded exercise submissions for about 300 students; preparation of the project regarding the implementation of the Rapid Spanning Tree Protocol (RSTP); answered questions in the online chat during lectures.

Spring 2021 **Algorithms and Probability**, *ETH Zurich*

Held weekly exercise sessions for about 20 students; graded exercise sheets; prepared quizzes and mini-tests covering the course content.

Fall 2020 **Algorithms and Data Structures**, *ETH Zurich*

Held weekly exercise sessions for about 20 students; graded exercise sheets; answered questions in the online chat during lectures.

Spring 2020 **Algorithms and Probability: intensive course**, *ETH Zurich*

Wrote a script containing the whole scope of the course; held a week of intensive sessions to assist the students in the exam preparation, providing them intuitive explanations of algorithms and tricks to perform well.

Spring 2020 **Parallel Programming**, *ETH Zurich*

Held weekly exercise sessions for about 20 students; helped the students implementing algorithms in Java; wrote the script for the course.

Fall 2019 **Algorithms and Data Structures**, *ETH Zurich*

Held weekly exercise sessions for about 20 students; graded exercise sheets; answered questions in the online chat during lectures.

2013-2021 **Mentored 30+ high-school students in Maths, Physics and Chemistry classes.**

## Talks

2023 Cote Alanna et al., Comparison of Confound Adjustment Methods in the Construction of Gene Co-expression Networks, Biostatistics Department JC, Harvard University.

2023 Siska Charlote et al., The Discordant Method: a novel approach for Differential Correlation, Biostatistics Department JC, Harvard University.

2023 Wei Liu et al., Inferring Gene Regulatory Networks Using the Improved Markov Blanket Discovery Algorithm , Biostatistics Department JC, Harvard University.

2022 Adam Bloniarz et al., Lasso Adjustments of Treatment Effect Estimated in Randomized Experiments, Biostatistics Department JC, Harvard University.

2022 Organizer and speaker at the IT Data Science Event by the Swiss Study Foundation.

2021 Francisco Ruiz et al., SHOPPER: a probabilistic model of consumer choice with substitutes and complements, *Advanced Topics in Machine Learning and Data Science Seminar*, ETH Zurich.

2019 Naeemul Hassan et al., Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster, *Seminar on Media Innovation* at ETH Zurich.

2018 Presented my high school project Random number generators and their applications in computer science with the Monte Carlo method at the Taiwan International Science Fair.

2017 Presented my high school project Random number generators and their applications in computer science with the Monte Carlo method at the National Competition by Schweizer Jugend Forscht.

## Graduate Coursework

| | |
|---|---|
| Harvard University | Reproducible Data Science (BST 270). |
| ETH Zurich | Advanced Machine Learning; Algorithms Lab; Big Data; Computational Statistics; Data Analytics for Non-Life Insurance Pricing; Financial Economics; High Dimensional Statistics; Information Theory; Machine Perception; Optimization for Data Science; Reinsurance Analytics; Statistical Modeling. |

## Undergraduate Coursework

| | |
|---|---|
| ETH Zurich | Algorithms and Data Structures; Algorithms and Probability; Algorithms, Probability and Computing; Analysis I-II; Complex Analysis; Computer Networks; Computer Systems; Data Modelling and Databases; Design of Digital Circuits; Discovering Management; Discrete Mathematics; Functional Programming and Formal Methods; Introduction to Machine Learning; Information Security; Introduction to Neuroinformatics; Introduction to Programming; Linear Algebra; Numerical Methods for CSE; Parallel Programming; Probability and Statistics; Systems Programming and Computer Architecture; Theoretical Computer Science; Visual Computing. |

## Technical Skills

| | |
|---|---|
| Languages: | Python, Java, R, C++ |
| Frameworks: | PyTorch (Lightning), Scikit-learn, Pandas, Numpy, Eigen, BGL, CGAL. |
| Algorithms: | Knowledge of Dynamic Programming, Graph Algorithms (Shortest Paths, Minimum Spanning Trees, Matching, Coloring, Flows in Networks, Min Cuts, ...), Geometric Algorithms (Delaunay Triangulation, Convex Hulls, Smallest Enclosing Circle, Point Location, Voronoi Diagrams, ...), Linear Programming, skilled in designing Randomized Algorithms, ... |
| ML: | Supervised Learning (Linear Regression, SVMs, Kernelized methods, MLPs, Decision Trees, KNN, Boosting, Bagging, ...); Unsupervised Learning (Clustering, Dimensionality Reduction); Deep learning (CNNs, GNNs, RNNs, Transformers); Recommender Systems. |
| Big Data: | SQL, Hadoop, Spark, MongoDB, MapReduce, JSONiq, Neo4j, Kafka. |
| Maths: | Solid foundations in linear algebra, probability, statistics, calculus, and optimization. |

## Swiss Study Foundation

The Swiss Study Foundation supports excellent students and postgraduates at universities and technical colleges who due to their personality, creativity and intellectual skills are in a position to contribute to science, business, culture and politics. The Foundation's mission is to promote young people who are willing and capable of assuming leading positions in all sectors of society. Since 2018 I am a proud member of the Swiss Study Foundation, that offered me highly valuable learning experiences complementary to my studies, together with diverse networking opportunities with people from various backgrounds. Here a list of selected events I attended and organized:

2022 **Planning for Health Development - Tanzania in Context**: local mission in Tanzania, visiting health facilities and presenting a new planning to the Swiss Embassy in Dar Es Salaam.

2022 **IT Data Science (co-organizer)**: introduction to fundamental topics in Data Science and Machine Learning (pipeline of a Data Science project, main ideas of supervised learning, bias-variance trade-off) and presentation of classical Deep Learning material (GANs, transfer learning, transformers).

2021 **Story telling camp:** presentation of different techniques to tell a story in an engaging way and prepared a "Pecha Kucha" presentation about my first solo travel experience.

2021 **NMUN:** negotiations about the Illicit Trade of Small Arms and Light Weapons in behalf of Madagascar at the NMUN event held in New York.

2020 **The Art of Debating:** introduction to the competitive form of British Parliamentary Debating. Theoretical sessions with the former European champion and a team tournament event.

2019 **IT Life Science:** introduction to genetics and talks about CRISPR/Cas9.

2019 **IT Personal Management:** discussions and ideas about life design.

## Languages

**English** (proficient); **German** (fluent); **Italian** (mother tongue); **French** (basic knowledge); **Spanish** (elementary knowledge).

## Others

2020 **Mentoring**

Welcome event for new ETH Student.

2016-2018 **Participant of the TV debating show *Linea Rossa***

- Linea Rossa was a popular TV show at RSI (the official television broadcaster in Ticino).
- Participated in weekly discussions and debates about various topics (politics, fake news, social media, university, dating apps, ...)