Ioannis Zachos

a.k.a. **Yannis**

Doctoral Researcher in Computational Statistics

Email
Website
LinkedIn

I am passionate about constructing Statistical Machine Learning (SML) algorithms for industry-relevant Research & Development. My research focuses on eliciting individual agent dynamics from aggregate data using Markov Bases, Markov Chain Monte Carlo, and Physics-informed ML.

Education

Oct. 2019 – Present

UNIVERSITY OF CAMBRIDGE

ARUP

Cambridge University (CU): *MRes* + *PhD* **fully funded** by Arup and EPSRC. *PhD* in Computational Statistics supervised by Prof. Mark Girolami, Prof. Theo Damoulas, Dr. Gerard Casey.

- Thesis: Probabilistic ecological inference in agent-based modelling: Beyond mean-field approximations.
- Collaborated with Arup's City Modelling Lab.

MRes in Future Infrastructure & Built Environment. **2 Commendation Letters** (**77% overall**).

- Thesis: Stochastic modelling of urban travel demand using Metropolis Hastings and Hamiltonian Monte Carlo (code).
- Courses: Computational Statistics and Machine Learning (83%).

Oct. 2015 – July 2018

WARWICK THE UNIVERSITY OF WARWICK

Warwick University (WU): *BSc* Data Science. **1st class honours**. Thesis supervised by Prof. Theo Damoulas.

- Thesis: Bayesian online change-point detection for time series segmentation and forecasting in non-stationary point processes (79%).
- Courses: Machine Learning (73%), Mathematical Statistics (79%), Linear Statistical Modelling (77%), Topics in Data Science (82%), Programming for Data Science (81%), Artificial Intelligence (72%).

Sep. 2013 - July 2015

Anatolia College: International Baccalaureate. **39/45** (**top 7% globally**).

- **Awarded 10,000 €** merit-based scholarship for academic excellence.
- Courses: Physics (7/7), Mathematics (6/7), Business Management (6/7), Extended essay on stock price forecasting using Statistics (35/36).

Work Experience

Sep. 2018 - July 2019

Cervest Ltd: Statistical Scientist.

- Research projects led:
  - Change-point detection on climate-volatile data generating processes.
  - Sequential multinomial classification for assessing environmental resilience.
  - Bayesian models for spatio-temporal image and sensor data fusion.
- Designed and developed data acquisition infrastructures using Python for use by the Data Science team.
- Engaged with clients and investors to facilitate science communication.

June - Aug. 2018

Eurobank Private Bank Luxembourg: Investment Advisory Intern.

- Designed, developed and deployed a web application for portfolio management.
- Derived optimal portfolios using efficient frontier theory with diversification and volatility constraints.

Research Experience

Oct. 2020 - Aug. 2021

Arup, CU: Model assessment of constitutive laws in traffic conservation laws. Research project supervised by Prof. Mark Girolami, Dr. Gerard Casey.

- Computed Bayes factors of constitutive laws embedded in traffic flow partial differential equations (PDEs) using thermodynamic integration (code).

Oct. 2019 - Jan. 2020 — National Highways, CU: Bayesian hydrological modelling of road rainfall run-off. Research project supervised by Prof. Mark Girolami (**80%**).

- Developed probabilistic hydrological model comparison and prediction framework using Sequential Monte Carlo (code).
- Collaborated with company executives to identify scope and present results.

Jan. - April 2018 — Kaggle competition, WU: Detection and Segmentation of nuclei from cell images. Machine Learning project (**84%**).

- Trained Multi-layer Perceptron and Convolutional Neural Network and compared against Watershed image segmentation.
- Performed data augmentation to achieve translation and rotation-invariance.

June - Aug. 2017 — Warwick University: Summarising large binary sequences for RNA editing. Research project supervised by Prof. Anastasia Papavasiliou.

- **Awarded 1000£** for outstanding performance in Mathematical Statistics to research methods of summarising large binary sequences.
- Utilised theory of rough paths to compute signatures of binary sequences.

Jan. - April 2017 — Deutsche Bank, WU: Anomaly detection of FTSE100 stocks. Team leader on Software Engineering group project.

- Developed a real-time machine learning platform that detected anomalies in one million daily transactions.
- Pitched our platform to company stakeholders.

Selected Publications

Under Review — *Generating Origin-Destination Matrices in Neural Spatial Interaction Models* (code). **Zachos**, Girolami, Damoulas.

Stat 2024 — *Table Inference for Combinatorial Origin-Destination Choices in Agent-based Population Synthesis* (code). **Zachos**, Damoulas, Girolami.

Skills

Programming: Python ($\approx 10^5$ lines), R ($\approx 10^4$ lines), Java, C, Matlab ($\approx 10^3$ lines each).

Libraries: `PyTorch`, `TensorFlow`, `numpy`, `PyMC3`, `OpenCV`, `sklearn`, `numba`.

Cloud: Amazon Web Services (S3, EC2), Google Cloud Platform.

Databases: MySQL, PostGIS/PostgreSQL.

GIS: QGIS, Google Earth Engine API, SentinelHub API, GDAL.

Miscellaneous: Unix Shell Scripting, Git, Data Version Control, LaTeX, R Shiny, Docker.

Languages: Greek (native), English (fluent).

Leadership Activities

Oct. 2023 - Present — Cambridge University Hellenic Society. Captain of Basketball team.

- **Secured 600 €** company sponsorship and led the team.

Sep. 2022 - Nov. 2022 — Annual Future Infrastructure & Built Environment Conference. Co-lead organiser of a team of 10.

- Attracted 50 attendees of which **95%** rated their experience as positive and **90%** said they would recommend this conference to colleagues.

Dec. 2019 - March 2020 — Judge Business School EnterpriseTECH. Team communicator in a team of 5.

- Developed business case for an air pollution prediction platform and pitched it to potential investors.

Oct. 2015 - July 2018 — Warwick University Statistics Dept. Student Representative, Mentor.

- Mentored students & liaised with staff to improve teaching quality and student support by collecting and discussing feedback.