RAFAEL CABRAL

Address: 339 Pasir Panjang Road #04-06 Singapore, 117368
Nationality/Age: Portuguese / 27yo
Mobile: +65 98852907
Email/Website: rafael@nus.edu.sg / rafaelcabral96.github.io/

STATISTICS AND COMPUTING
(RESEARCH AND DEVELOPMENT | COMPUTATIONAL AND APPLIED STATISTICS | PROGRAMMING)

Summary: I have a solid background in statistics and programming with R, Python, and C++. As a statistics researcher, I specialized in applied and computational statistics. My research revolved around building more robust and computationally efficient modeling frameworks for spatial and temporal data. I also did internships in the industry in data science, deep neural networks, and computer vision.

EDUCATION

Jun 2023 **Doctor of Philosophy in Statistics**
King Abdullah University of Science and Technology (KAUST), Saudi Arabia
Dissertation Title: **Criticism and robustification of latent Gaussian models**
GPA: 3.93/4 (Advisors: Profs. Haavard Rue and David Bolin)

Jun 2019 **Master of Science in Applied Mathematics**
Instituto Superior Técnico - University of Lisbon, Portugal
Dissertation Title: **Space-time trends and dependence of wind extremes in north-western Germany**
GPA: 3.87/4; Top degree in Portugal: the admissions GPA is 3.70/4

Jun 2017 **Bachelor of Science in Engineering Physics**
Instituto Superior Técnico - University of Lisbon, Portugal
GPA: 3.70/4; Top degree in Portugal: the admissions GPA is 3.78/4

RESEARCH EXPERIENCE

– Doctor of Philosophy in Statistics

Key Highlights –

- My research dealt with a generic class of statistical models (latent Gaussian models), which include most spatial and temporal models used in practice in fields such as climate science, disease mapping, and econometrics. Tools I am familiar with: Bayesian statistics and approximate inference (Markov Chain Monte Carlo, variational inference, and Laplace approximations), spatial and temporal models, and Gaussian and non-Gaussian stochastic processes.

- *Statistical assumptions are rarely met in practice.* I developed new frameworks for a) checking assumptions in general (Bayesian hierarchical) models and their impact on predictions and b) constructing and fitting more robust models.

- By combining Variational Bayes with Laplace approximations, we obtained algorithms that converged orders of magnitude faster than MCMC (**10-10000x**), depending on the data size, while maintaining reasonable accuracy.

- Built an R package where the previous workflow was implemented (also using C++): **ngvb package**

- The final models are more robust, less sensitive to outliers in the data, and more accurate at predicting extreme events, such as spikes in spatial data (for instance, hotspots in temperature data) or sudden jumps in time-series data. We had a 20% accuracy increase in predicting atmospheric pressure data.

- My advisors were awarded a **1-million-dollar grant** based on my research to continue this line of work. Papers published in prestigious journals:

  Robustness, model checking and latent Gaussian models,
  R. Cabral, D. Bolin, and H. Rue, 2023. **Journal of the Royal Statistical Society Series B (in revision).**

  Fitting latent non-Gaussian models using variational Bayes and Laplace approximations
  R. Cabral, D. Bolin, and H. Rue, 2022. **Journal of the American Statistical Association.**

  Controlling the flexibility of non-Gaussian processes through shrinkage priors,
  R. Cabral, D. Bolin, and H. Rue, 2022. **Bayesian Analysis.**

– Master of Science in Applied Mathematics

Key Highlights –

- Key courses: Mathematical statistics, probability theory, stochastic processes, network science, and machine learning.
- **Bitcoin Transaction Network Analysis**: a project for a course where we analyzed the topology of the Bitcoin transaction network and the existence of sporadic bursts of transactions.
- Developed new methods in extreme value theory for evaluating temporal trends and spatial homogeneity in extreme events. Modelling and prediction of precipitation and wind extremes in Germany. Paper:

  Space-time trends and dependence of precipitation extremes in north-western Germany,
  R. Cabral, A. Ferreira, and P. Friederichs, 2019. **Environmetrics.**

- Developed a price-formation model with a population consisting of a finite number of agents storing and trading a commodity. We formulate our problem as an N-player dynamic game with a market-clearing condition. Paper:

  A price model with finitely many agents,
  A. Alharbi, T. Barkaryan, R. Cabral, et al, 2019. **Bulletin of the Portuguese Mathematical Society.**

– Bachelor of Science in Engineering Physics

Key Highlights –

- Proficiency in foundational principles of mathematics and physics. Key courses: Calculus, linear algebra, differential equations, statistical physics, electromagnetism, and quantum mechanics.
- Practice in the programming languages C and C++ acquired through coursework in programming, computational physics, and computational mathematics. Successfully completed projects involving numerical solutions of partial differential equations to simulate gravitational systems and Monte Carlo simulations of radioactive decay.
- Developed an **Android App** for a science communication course called "Who wants to be a Physicist?" using Java and Android Studio. A visually dazzling quiz app where you can answer randomly generated physics quizzes according to chosen difficulties or themes.

WORK EXPERIENCE

– National University of Singapore

As a Research Fellow at the **Division of Biomedical Data Science** (2023-2024)

Key Highlights –

- I am deriving new clustering algorithms based on Bayesian (non-parametric) principles. These algorithms allow for better uncertainty number of clusters is unknown) and prediction quality while maintaining a speed comparable to k-means.
- Clustering a large dataset of ancient coins in SE Asia: I used computer vision tools (CNNs, SVM, random forests, etc.) to extract features from the images and then apply the clustering algorithm.
- **Accuracy of 96%,** never-before-seen in the field of numismatics since the differences between coins are very subtle.
- **Consultancy work**: several applied statistical analyses for the Singapore Institute for Clinical Sciences, including functional data analysis in a study that attempts to link brain health to nutrition and causal inference and mediation analysis in clinical trials.

– King Abdullah University of Science and Technology

As Teaching Assistant (Fall Semesters in 2020, 2021 and 2022)
As Student Ambassador (2020-2023)

Key Highlights –

As Teaching Assistant of the course "Stochastic Processes" (Fall Semesters in 2020, 2021 and 2022)

- Mandatory M.Sc./Ph.D. course in Statistics and Electrical Engineering with 45 attending students.
- Designed course projects and weekly homework and graded exams and projects.
- Delivered weekly lectures where I solved the homework and answered students' questions.

As Student Ambassador (2020-2023)

- Promoted the Statistics program abroad in conferences and universities and helped organize recruitment events.

INTERNSHIPS AND RESEARCH VISITS

– **Research Visit** at the **RIKEN** center for the **Advanced Intelligence Project**, Tokyo, Japan, 2024

- Invited by **Emtiyaz Khan** of the Approximate Bayesian Inference Team: utilizing Bayesian statistics to improve and explain AI models.

– **Summer School: Statistics for Point Patterns in Space and Beyond**, Aalborg, Denmark, 2022

– **Internship on Optimization and Machine Learning** in KAUST, Saudi Arabia, 2019

- Implemented and improved a particle-swarm optimization algorithm. Developed the Python package **mlswarm**.
- Application to non-convex function mi to non-convex function optimization and machine learning problems: requires no gradient computations and about **10x-50x fewer function evaluations** compared to Nelder-Mead, Differential Evolution, Mesh Search, and Simulated Annealing.

– **Researcher at Hospital da Luz Learning Health**, Portugal, 2018

- Used deep convolutional neural network to segment tumors in 3D brain images using Keras and TensorFlow.
- Won an Enterprise Challenge competition at my university on how AI can aid the medical industry.

– **Spring Week at BNP Paribas**, Portugal, 2018

- Workshops (Alteryx and Tableau), work shadowing, informative presentations and case studies

– **Data Scientist summer internship at Odysai**, Portugal, 2018

- Startup developing a product that automatically extracts information from documents using machine learning.
- Data processing tasks using Python and stress testing the client-server architecture and ML product. I also designed real-time analytics dashboards so that stakeholders could make data-driven decisions on user behavior and the performance of the ML algorithm.

HONORS AND AWARDS

- **Al-Kindi Statistics Research Student Award**, a research award for top Statistics Ph.D. students at KAUST (2022)
- **McKinsey & Company Forward program** (2022)
- **Dean's List Award** during the Ph.D. at KAUST (2022)
- **Ph.D. progress rated as Outstanding (A+)** by the KAUST evaluation committee (2020-2022)
- **Diploma of academic merit** in the M.Sc. in Applied Mathematics (2019)
- **Research fellowship grant** for the development of research work in extreme value theory (2018)
- **Diploma of academic merit** in the B.Sc. in Engineering Physics (2017)
- **Asteroid Discovery Award** by the International Astronomical Search Collaboration. Confirmed asteroid by the Minor Planet Center of the University of Harvard with name 2013 EZ7 (2013)

SKILLS

Software	Mathematica, Matlab, MS Office Suite
Computer Programming	R, Python, C, C++, LATEX, Git, and to a lesser extent Java, Javascript and SQL
Packages	Pandas, Keras, Tensorflow, BUGS, Stan, R-INLA
Online courses	**Financial markets** from Yale University, **SQL for Data Science** from UC Davis