Miguel Raúl Pebes Trujillo

First name: Miguel, Middle name: Raúl, Last name: Pebes Trujillo

Contact information

Email: mpebes@ntu.edu.sg

Office: Nanyang Technological University, Research Techno Plaza (RTP), X-Frontiers Block, Level 6, 50 Nanyang Ave, Singapore 639798

Research interests

Development of statistical models and inference algorithms for understanding and controllability of real-world systems. I have worked in various domains including telecommunications, fraud and pattern-change detection, mobility analysis, risk management, hierarchy and network inference, and currently remote sensing and precision agriculture. Technical Core: Bayesian Modeling, MCMC methods, Machine Learning, and the Philosophy of Inductive Reasoning.

Education

Ph.D. in Statistical Science; Ph.D. Minor in Complex Networks & Systems; Indiana University Bloomington; USA — 2023
Co-Advisors: Daniel Manrique-Vallier, Andrew J. Womack

M.S. in Statistics; Pontifical Catholic University of Peru — 2015

B.S. in Computer Science; Pontifical Catholic University of Peru — 2008

Academic Positions

- **Research Fellow in Artificial Intelligence and Data Science** — 2023-2025
  School of Materials Science and Engineering, Nanyang Technological University, Singapore

  Singapore imports ∼90% of its food and aims to meet 30% of nutritional needs locally by 2030. I am working on models to optimize crop yield, nutrition, and growth cycles to make urban farming economically sustainable. I have developed Bayesian reinforcement learning models that control the daily irrigation of experimental plants in a greenhouse in northwest Singapore. I am also working on the discovery of new remote sensing indexes that predict crop traits, enabling translatable, affordable AI-based precision agriculture systems. Project funded by the National Research Foundation (NRF), Singapore. Lead PI's: Kee Woei Ng and Matan Gavish.

  I am currently co-leading the elaboration of a research grant proposal for ∼S$2.3 million, to be submitted to NRF - Central Gap Fund, a research funding initiative for translational research.

- **Visiting Assistant Professor** — 2022-2023
  Department of Statistics, Indiana University Bloomington, USA

Stat-S 520: Introduction to Statistics	FA2022
Stat-S 611: Applied Statistical Computing	SP2023
Stat-S 301: Applied Statistical Methods for Business	SP2023

  My students' majors were varied including M.S. in Data Science, Ph.D. and M.S. in Computer Science, Ph.D. in Information Science, Ph.D. in Economics, and M.A. in International Studies.

- **Teaching Assistant** — 2016-2022
  Department of Statistics, Indiana University Bloomington, USA

Stat-S 426/626: Bayesian Theory and Data Analysis	FA2021, FA2020, F2019, S2019
Stat-S 611: Applied Statistical Computing	SP2022, SP2021, SP2020, FA2018, FA2016
Stat-S 431: Applied Linear Models I	FA2021, FA2020
Stat-S 352: Data Modeling and Inference	SP2021
Stat-S 670: Exploratory Data Analysis	SP2018, SP2017
Stat-S 350/520: Introduction to Statistics	SU2020, SU2019
Stat-S 303: Statistics for the Life Sciences	FA2016

- **Research Assistant** 2018-2020
  Department of Statistics, Indiana University Bloomington, United States of America

  Project *Bayesian approaches to financial auditing*: Developed a risk-assessment framework using graphical models and natural language processing to estimate the probability of material misstatement in ledgers from anomaly patterns in structured and unstructured audit data. Funded by Deloitte & Touche. PI's: Daniel Manrique-Vallier and Andrew J. Womack.

Peer Review Service

- **Scientific Committee Member** 2025
  IEEE/GRSS International Geoscience and Remote Sensing Symposium (IGARSS)
  Topic codes: T/D: Data Analysis, T/A: AI and Big Data, CCS.45: Crop Growth Models and RS

- **Journal Reviewer** 2024
  IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
  Under keywords: Statistics, Intelligent Systems, Artificial Intelligence

Research Papers

1. **Pebes-Trujillo, Miguel**; Shenhar, Itamar; Harikumar, Aravind; Herrmann, Ittai; Moshelion, Menachem; Kee Woei, Ng; Gavish, Matan (2024). Ranking Learning for Experiment Evaluation 2024
2. **Pebes-Trujillo, Miguel**; Shenhar, Itamar; Harikumar, Aravind; Herrmann, Ittai; Moshelion, Menachem; Kee Woei, Ng; Gavish, Matan (2024). Hierarchical Change-Point Detection (in progress) 2024
4. Harikumar, Aravind; Shenhar, Itamar; **Pebes-Trujillo, Miguel**; Lin, Qin; Moshelion, Menachem; He, Jie; Ng, Kee Woei; Gavish, Matan; Herrmann, Ittai (2024). Harnessing Smartphone RGB Imagery and LiDAR Point Cloud for Enhanced Leaf Nitrogen and Shoot Biomass Assessment - Chinese Spinach as a Case Study 2024
5. Harikumar, Aravind; Shenhar, Itamar; **Pebes-Trujillo, Miguel**; Lin, Qin; Moshelion, Menachem; He, Jie; Ng, Kee Woei; Gavish, Matan; Herrmann, Ittai (2024). Smartphone LiDAR Reliably Estimates Leaf Nitrogen Concentration and Shoot Biomass on Leafy Vegetable Crops. IGARSS 2024 2024
6. **Pebes-Trujillo, Miguel**; Manrique-Vallier, Daniel; Ahn, Yong-Yeol; Womack, Andrew J. (2022). Inference of Hidden Hierarchies from Observable Networks (in progress) 2022
7. **Pebes-Trujillo, Miguel**; Manrique-Vallier, Daniel (2022). Burst Detection in Call Trains for Identifying Fraud in Telecommunications (in progress) 2022

Theses

1. **Pebes-Trujillo, Miguel** (2023). Inference of Hidden Hierarchies From Observable Networks 2023
   Ph.D. dissertation. Co-advisors: Daniel Manrique-Vallier, Andrew J. Womack
2. **Pebes-Trujillo, Miguel** (2014). An Application of Discrete-Time Survival Models to Analyze Student Dropouts at a Private University in Peru 2014
   M.S. thesis. Advisor: Giancarlo Sal y Rosas

Presentations

1. **Poster** at the *Global Young Scientists Summit 2025*, NRF Singapore (Jan 6-10) 2025
   T: Crop Trait Optimization via Precision Agriculture: Bridging Machine Learning and Food Security
2. **Invited Speaker** at *Association of Women Entrepreneurs of Peru*, Lima, Peru (Oct 17) 2024
   Title: AI for Business Sustainability
3. **Contributed Speaker** at conference *Networks 2021* (Jul 6) 2021
   Title: Hidden-Hierarchy Reconstruction from Bipartite Networks
4. **Poster** at the *Symposium on Data Science and Statistics 2019*. Bellevue, Washington (May 30) 2019
   Title: Burst Detection in Call Trains for Identifying Fraud in Telecommunications

Relevant Conference Attendance

1. Global Young Scientists Summit (GYSS) 2025. World meeting of young scientists with twenty Nobel Laureates and Eminent Scientists. Stephen Riady Centre, National University of Singapore (January, 2025). Attendance by institutional nomination only. 2025
2. Bayesian Nonparametrics (BNP) Networking Workshop 2024 (Topic: Interpretable Inference via Principled BNP Approaches in Biomedical Research and Beyond). Institute of Mathematical Sciences, National University of Singapore. Co-chairs: Maria de Iorio, Peter Müeller, Igor Pruenster 2024
3. AI4Science and Nobel Turing Challenge Initiative Conference. Organizers' affiliations: NRF; Agency for Science, Technology and Research (A*STAR); NTU; and the National University of Singapore (NUS) from Singapore; and Sony AI, The Systems Biology Institute, and RIKEN, from Japan. 2024

Non-degree graduate training

Ph.D. Visiting Student. Department of Network and Data Science. Central European University 2021
International Program in Data Analysis. Entrepreneurship Development Institute, India 2013
Summer Program in Statistics. University of Sao Paulo and Federal University of Sao Carlos, Brazil 2013
Specialization Program in Predictive Analysis and Data Mining. Knowledge and Systems Peru, Lima 2010
Diploma in Business Intelligence. Peruvian University of Applied Science, Lima, Peru 2010

Professional Experience

Before my Ph.D., I spent eight years in Data Science, Analytics, Engineering, and Translation roles at a telecom company, developing data mining and machine learning models on large-scale datasets including phone calls, location data, text messages, and browsing records. These models were applied to strategic planning, CRM, risk management, fraud detection, predictive modeling and automation.

- Positions at Entel Peru S.A. Telecommunications; Lima, Peru 2015-2016

  Risk Management Planning Coordinator (senior data science role)

  Provided consulting to upper management by developing predictive models for customer attrition, "loan" defaults, and user segmentation; monitoring credit scoring models' calibration; assessing proposed credit-policy risks; and leading rapid data analyses for decision making on risk and fraud prevention. (Oct 1, 2015 - Jul 1, 2016)

  Senior Process Intelligence Coordinator (senior data science role)

  Led a team of computer scientists on projects for the Finance Central Department, focusing on quantitative analysis, data visualization, and unstructured data insights. Automated predictive models, developed KPI dashboards, and managed databases. Key projects included phone call traffic and mobility analysis, user segmentation, debt forecasting, billing complaint prediction, and criminal activity detection. (Jul 1, 2015 - Sep 30, 2015)

- Positions at Nextel del Peru S.A., Telecommunications (subsidiary of NII Holdings Inc.) 2009-2015

Process Intelligence Coordinator	Jun 1, 2014 - Jun 30, 2015
Senior Processes Intelligence Analyst	Apr 1, 2011 - May 31, 2014
Analyst of Financial Processes Management	Jan 1, 2011 - Mar 31, 2011
Junior Analyst of Financial Processes Management	May 1, 2009 - Dec 31, 2010
Intern of Financial Processes Management	May 1, 2008 - Apr 30, 2009

Software consulting

Consultant for the Forest Resource Monitoring Agency (OSINFOR), Peruvian Government 2016-2017
Project: Implementation of a geo-referenced information system for forest supervisions.
For our software, *GeoSupervisor*, OSINFOR was awarded the Prize for Good Practices in Public Management 2018.

Advanced statistical and computational skills

Bayesian statistics: Bayesian modeling and computational posterior inference. Markov Chain Monte Carlo (MCMC) (including Gibbs samplers, Metropolis-Hasting algorithms, adaptive strategies), diagnostics, model selection, and graphical models.
Machine learning: Supervised and unsupervised methods, generalized linear models (GLMs), hidden Markov models (HMMs), mixture models, graph and DAG inference, etc. Bayesian reinforcement learning and Thompson sampling approaches.
Data mining and pattern recognition: Feature selection, outlier detection, change point detection, rule system inference, mining of linear and non-linear implicit relations, etc.
Numerical methods: RNGs, Expectation-Maximization (EM), Newton-Raphson algorithms, Monte Carlo optimization, Bootstrap, heuristic methods (e.g. grid search, random walks strategies, simulated annealing, and genetic algorithms).
Classic statistics: Probability and large sample theory, point and interval estimation, hypothesis testing (e.g. ANOVA), multivariate analysis, time series, etc.
Scientific programming: R, C/C++, Python. I also use these tools: Bash, SSH, Tmux, Emacs.
Programming paradigms: Sequential, object-oriented, and dynamic programming, automation and batch processing, memory management and bit-level programming, APIs, AWS data workflows.
Operating Systems: Mac OS, Linux, Windows.

Additional information

Memberships:	International Society for Bayesian Analysis (ISBA), Sections: Bayesian Nonparametrics, Bayesian Computation, Environmental Sciences.
Languages:	English, fluent in all areas. Spanish, native speaker.

Last update: February, 2025. [1]

[1] Available upon request: (1) Degree certificates. (2) Teaching and Research statements. (3) Syllabi of the courses I've taught. (4) References. (5) Honors and Awards in academia and private sector. (6) Transcripts.