Daniel **long**

Statistics Ph.D Candidate

(+1) 510-816-8686 | daniong@umich.edu | danieliong.github.io | danieliong | danieliong | US Citizen

Education

University of Michigan, Ann Arbor (UM) — *Ann Arbor, MI*

Ph.D in Statistics — *08/2017 – **05/2023 (expected)***

- *Rackham Merit Fellow*

University of California, Davis (UCD) — *Davis, CA*

B.S. Statistics (high honors), B.A. Economics (honors), Chinese Minor — *10/2013 – 06/2017*

- *Outstanding Academic Performance Citation (Dept. of Statistics)*, Dean's List

Experience

Orbital Insight — *Palo Alto, CA*

Data Scientist Intern — *05/2022 – 08/2022*

- Developed algorithm based on convolutional neural networks and other computer vision/geospatial analytics methods to detect rare GNSS interference events in geolocation (AIS, ADS-B) data as a key deliverable for a *Department of Defense contract*.
- Trained convolutional neural networks on generated synthetic data using PyTorch on AWS EC2 instances.
- Worked with software engineers to productionize algorithm by integrating it onto the company's flagship GO platform.

NASA Goddard Space Flight Center (Solar Physics Laboratory) — *Virtual*

Research Intern [final presentation slides] — *06/2021 – 08/2021*

- Collaborated with solar physicists with minimal statistical training to develop new methods/metrics for evaluating an empirical solar wind model.
- Extended dynamic time warping to account for domain-specific issues when using it for solar wind model evaluation.
- Created *web app* in Python using Dash, Plotly to visualize dynamic time warping for model evaluation.

Department of Statistics, UM — *Ann Arbor, MI*

Graduate Student Instructor — *09/2018 – 04/2020*

- **Courses**: Intro. to Statistics & Data Analysis (undergraduate), *Bayesian Data Analysis* (undergraduate), Bayesian Modeling & Computation (graduate)

Projects

Explainable machine learning for space weather forecasting

Solar Storms & Terrestial Impacts (SOLSTICE) Center, UM [project website] — *02/2021 – 09/2021*

- Trained gradient boosted trees (XGBoost) to predict high-resolution geomagnetic index several hours ahead in Python, resulting in a 10% lower RMSE compared to the best existing forecasting methods in the space weather literature.
- Collaborated with space scientists to explain predictions using explainable ML methods (SHAP), leading to novel insights about underlying physics.
- Created *web app* in Python using Dash, Plotly to visualize results; Presented this work to 20+ space scientists at invited seminar talk.
- Published *first-author paper* in AGU Space Weather journal.

Modeling heterogenous causal mechanisms in epidemiology with observational data

Department of Statistics, UM [project website] — *05/2019 – 08/2020*

- Developed novel probabilistic clustering method to model causal mechanisms between HDL cholesterol and coronary heart disease.
- Implemented Monte-Carlo EM algorithm in R/C++ to perform statistical inference (parameter estimation, confidence intervals, model selection).
- Submitted *first-author paper* to Annals of Applied Statistics; Presented work to 100+ epidemiologists/statisticians at several conferences/seminars.
- Developed and wrote documentation for *R package (MR-PATH)*.

NOAA Forecasting Competition: Modeling the Geomagnetic Field

SOLSTICE Center, UM [Competition results] — *01/2021 – 02/2021*

- Ranked top 5% (32/623) in competiton hosted by NOAA (1st place prize: $15,000) to forecast a geomagnetic index under operationally viable constraints.
- Collaborated with domain experts to write *custom Scikit-learn transformers* to clean/preprocess real-time solar wind data with > 8mil. observations.
- Trained various models including gradient boosted trees, feed-forward/long-short term memory neural networks in Python.

Variational spatial Gaussian processes for space weather forecasting [Work in progress]

Department of Statistics, UM — *09/2022 - Present*

- Extended variational inference (VI) algorithm to perform inference for sparse spatial Gaussian processes with novel heavy-tailed likelihood.
- Implemented novel VI algorithm in PyTorch/GPyTorch to train models for forecasting geomagnetic perturbations.

Skills

Programming Languages	Python, R/Rcpp, Julia, C++, SQL (Postgres)
Data Science Tools	Numpy, Pandas, Scikit-learn, XGBoost, PyTorch(-Lightning), Stan, ggplot, Matplotlib, Plotly, Dash
Data Science Methods	Bayesian/probabilistic modeling, time series forecasting, causal inference, statistical computing, deep learning
Computing Tools	AWS (EC2, S3), Shell scripting, Linux (Ubuntu, Arch), High Performance Computing (Slurm)
Collaboration tools	Version control (Git), Confluence, JIRA