Daniel **long**

Statistics Ph.D Candidate

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

Education

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

Ph.D in Statistics — *08/2017 – **12/2022 (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

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

Research Intern (Supervisor: Charles N. Arge) [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)
- Prepared lectures on advanced topics in **Bayesian modeling** (mixture models, model checking, etc).
- Advised students on extensive projects in applied Bayesian analysis using Stan.
- Taught weekly labs on basic statistical concepts, held weekly office hours to answer homework questions.

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.
- Wrote *first-author paper* with domain experts and submitted it to the 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.

Skills

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

Certifications

Master SQL For Data Science	Udemy [*Certificate*]
Algorithmic Toolbox	Coursera [*Certificate*]