

Yunan Gao

Houston, TX 77006

yg54vg8hu_a87@indeedemail.com

+1 346 379 1960

I am currently a fourth-year PhD student in Statistics at Rice University. Generally speaking, my research is focused on Bayesian statistical modeling with an emphasis on high-dimensional/functional data analysis and its applications.

I am also interested in the combination of modern statistical tools and real world industry problems, given my undergraduate training in Materials science and engineering. I am actively looking for data science/analytics internship opportunities for Summer 2022.

Willing to relocate: Anywhere

Work Experience

Graduate Research Assistant

Rice University - Houston, TX

June 2018 to Present

- Created a model to map the relationship between a scalar response and a high dimensional trajectory-like predictor using Bayesian functional regression and smoothing splines.
- Used R to create a MCMC algorithm that efficiently produce posterior estimates of the association surface.
- Developed new decision analysis tools to extract the critical windows in the high dimensional trajectory-like predictors that are most important to predict the scalar response.
- Collaborated with the researchers in Children's Environmental Health Initiative (CEHI) to investigate the relationship between the educational outcome and prenatal (daily) PM2.5 exposure level among a large cohort of children in North Carolina (>100,000).
- Handled large survey data and performed extensive exploratory data analysis, feature selection, statistical modeling, and decision analysis using R. Concluded that the 3rd trimester is the most important window having negative impact on educational outcomes.
- Collaborated with faculty member at Rice University to write, revise, and submit paper for publication.

Graduate Teaching Assistant

Rice University - Houston, TX

August 2018 to May 2020

Courses I worked for:

- Elementary Applied Statistics (STAT 280)
- Bayesian Statistics (STAT 525)
- Advanced Statistical Methods (STAT 411/ STAT 616)

Work I did as a TA:

- Provided office hours for large groups of undergraduate/graduate students.
- Organized coding labs and helped students with using R.
- Produced solutions for homework and exams.

- Summarized information from students homework and updated them to the instructors.

Education

Ph.D. in Statistics Houston

Rice University - Houston, TX

September 2018 to Present

Bachelor of Science in Materials Science and Engineering Beijing

Tsinghua University - Beijing

September 2014 to July 2018

Skills

- R (4 years)
- Python
- SAS
- Latex
- Data Analysis
- Machine Learning
- Research (3 years)
- Statistics
- Regression Analysis (4 years)
- Data Science
- Statistical analysis (4 years)
- Data Visualization
- Time Series Analysis (3 years)
- Functional Data Analysis (2 years)
- Bayesian Inference (4 years)
- Simulation Design (4 years)
- Computation Statistics
- Higher education teaching (2 years)
- GitHub
- Statistical software (4 years)
- data mining (4 years)
- predictive modeling (4 years)

Languages

- English - Fluent
- Chinese - Expert

Links

<https://github.com/YunanGao>

Assessments

Analyzing data — Highly Proficient

March 2022

Interpreting and producing graphs, identifying trends, and drawing justifiable conclusions from data

Full results: [Highly Proficient](#)

Indeed Assessments provides skills tests that are not indicative of a license or certification, or continued development in any professional field.

Publications

Bayesian adaptive and interpretable functional regression for exposure profiles

<https://arxiv.org/abs/2203.00784>

March 2022

Pollutant exposures during gestation are a known adverse factor for birth and health outcomes. However, the links between prenatal air pollution exposure and educational outcomes are less clear, in particular the critical window of susceptibility during pregnancy. Using a large cohort of students in North Carolina, we study prenatal PM_{2.5} exposures recorded at near-continuous resolutions and linked to 4th end-of-grade reading scores.

- We develop a locally-adaptive Bayesian regression model for scalar responses with functional and scalar predictors.
- The model is designed to capture both smooth and rapidly-changing features in the regression surface and accompanied by a highly scalable Gibbs sampler for fully Bayesian inference on large datasets
- We describe broad limitations with the interpretability of scalar-on-function regression models, and introduce new decision analysis tools to guide the model interpretation.

Using these methods, we identify a period within the third trimester as the critical window of susceptibility to PM_{2.5} exposure.

R code for this new method is available in here: <https://github.com/YunanGao/BASOFR>