

# Siyu Zhou

Department of Biostatistics and Bioinformatics  
Rollins School of Public Health  
Emory University, Atlanta, GA

✉ [siyu.zhou@emory.edu](mailto:siyu.zhou@emory.edu)  
🌐 <https://syzhou5.github.io/syzhou5/>

## EDUCATION

- 2017.08 - 2022.07 **University of Pittsburgh**  
Ph.D. in Statistics  
Dissertation: *Random Forests and Regularization*  
Supervisor: Dr. Lucas Mentch
- 2014.09 - 2016.09 **The Hong Kong University of Science and Technology**  
M.Phil. in Mathematics  
Supervisor: Dr. Man-Yu Wong
- 2010.09 - 2014.05 **The Hong Kong University of Science and Technology**  
B.S. in Mathematics (First Honours)  
Minor in Actuarial Mathematics

## EMPLOYMENT

- 2022.07 - present **Postdoctoral Fellow**  
Department of Biostatistics and Bioinformatics,  
Rollins School of Public Health, Emory University  
Supervisor: Dr. Limin Peng

## RESEARCH INTERESTS

Machine learning, variable selection, quantile regression, survival analysis, tensor regression, applications in biomedical studies

## PUBLICATIONS AND MANUSCRIPTS

\* indicates equal contributions

- [1] **Zhou, S.** and Peng, L. “Approximate global censored quantile random forest,” In preparation.
- [2] **Zhou, S.** and Peng, L. “Global quantile learning with censored data based on random forest,” Submitted.
- [3] Wallace, M. L., Mentch, L., Wheeler, B. J., Tapia, A. L., Richards, M., **Zhou, S.**, Yi, L., Redline, S., and Buysse, D. J. (2023). “Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction,” *BMC medical research methodology*, 23 (1), 144.
- [4] **Zhou, S.\*** and Mentch, L.\* (2023). “Trees, forests, chickens, and eggs: when and why to prune trees in a random forest,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 16 (1), 45–64.
- [5] Mentch, L.\* and **Zhou, S.\*** (2022). “Getting better from worse: Augmented bagging and a cautionary tale of variable importance,” *Journal of Machine Learning Research*, 23 (224), 1–32.

- [6] Hooker, G., Mentch, L., and **Zhou, S.** (2021). “Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance,” *Statistics and Computing*, 31, 1–16.
- [7] Mentch, L.\* and **Zhou, S.\*** (2020). “Randomization as regularization: A degrees of freedom explanation for random forest success,” *Journal of Machine Learning Research*, 21 (171), 1–36.

## CONFERENCES AND PRESENTATIONS

### Talks

- 2024 “Why Random Forests Work and Why That’s a Problem.” Conference on Statistical Learning and Data Science, Newport Beach, CA. Nov 5 - 8.
- 2024 “Trees, Forests, Chickens, and Eggs: When and Why to Prune Trees in a Random Forest.” Joint Statistical Meeting, Portland, OR. Aug 3 - 8.
- 2022 “Regularization on Ensembles of Tree and Variable Importance.” ICSA Applied Statistics Symposium, University of Florida, Gainesville, FL. Jun 19 - 22.
- 2021 “Random Forests: Why They Work and Why That’s a Problem.” Joint Statistical Meeting, Virtual. Aug 3 - 8.
- 2021 “Augmented Bagging as an Alternative to Random Forests and Implications on Variable Importance.” Symposium on Data Science and Statistics, Virtual. Jun 2 - 4.
- 2020 “Augmented Bagging as an Alternative to Random Forests” Joint Statistical Meeting, Virtual. Aug 2 - 6.
- 2020 “Explaining the Practical Success of Random Forests.” Symposium on Data Science and Statistics, Virtual. Jun 3 - 5.

### Posters

- 2022 “Getting Better from Worse: Augmented Bagging and A Cautionary Tale of Variable Importance.” MDS-Rely Spring Meeting, Pittsburgh, PA. Apr 12 - 13.
- 2021 “Getting Better from Worse: Augmented Bagging and A Cautionary Tale of Variable Importance.” ASA Pittsburgh Chapter Spring Banquet, Virtual. Apr 23.
- 2019 “Regularization through Randomization: A Degrees-of-Freedom Explanation for Random Forest Success.” ASA Pittsburgh Chapter Spring Banquet, Pittsburgh, PA. Apr 16.

## AWARDS AND HONORS

- 2022 **ASA Student of the Year**, ASA, Pittsburgh Chapter
- 2021 - 2022 **Andrew W. Mellon Predoctoral Fellowship**, University of Pittsburgh
- 2021 **Outstanding Senior Graduate Student**, University of Pittsburgh
- 2016 **Best Mathematics Teaching Assistant Award**, The Hong Kong University of Science and Technology
- 2014 - 2016 **Asian Future Leaders Scholarship**, The Hong Kong University of Science and Technology and Bai Xian Asian Institute
- 2010 - 2014 **HKUST Mathematics Department Scholarship**, The Hong Kong University of Science and Technology

2010 - 2011, 2014    **Dean's List**, The Hong Kong University of Science and Technology

## TEACHING

G: Graduate level, UG: Undergraduate level.

### University of Pittsburgh

Instructor:

Stat 0800: Statistics in the Modern World (UG), SUM 2021

Stat 1000: Applied Statistical Methods (UG), SUM 2020

Teaching Assistant:

Stat 1000: Applied Statistical Methods (UG), SP 2018, SUM 2019, FA 2019, FA 2021

Stat 1100: Statistics and Probability for Business Management (UG), FA 2018, SP 2020

Stat 1221: Applied Regression (UG), SP 2019

Stat 1361: Statistical Learning and Data Science(UG), SP 2022

Stat 2270: Data Mining (G), FA 2019, FA 2021

Stat 2360: Statistical Learning and Data Science(G), SP 2022

### The Hong Kong University of Science and Technology

Teaching Assistant:

Math 3423: Statistical Inference (UG), FA 2015, FA 2016

Math 3424: Regression Analysis (UG), SP 2015, SP 2016

## SERVICE

2024.10                    Poster reviewer, Georgia Statistics Day

2023 - 2024                Survival and Longitudinal Data Analysis Study Group organizer, Department of Biostatistic and Bioinformatics, Emory University

2021 - 2022                Teaching Assistants/Fellows mentor, Department of Statistics, University of Pittsburgh

## SKILLS

Programming skills: R, SAS, MATLAB, Python,  $\text{\LaTeX}$ , C++

Languages: Mandarin (Native), English (Fluent)