

Soumyakanti Pan

Curriculum Vitæ

Padelford Hall C024
University of Washington
Box 354322
Seattle, WA 98195-4322
✉ span18@uw.edu
🌐 span-18.github.io
in [soumyakanti-pan](https://soumyakanti-pan.github.io)
🔗 [SPan-18](#)



Education

- 2021–2025 **PhD**, *Biostatistics*, University of California Los Angeles, Los Angeles, USA
2019–2021 **M.Stat**, *Statistics*, Indian Statistical Institute, Kolkata, India
2016–2019 **B.Stat (Honours)**, *Statistics*, Indian Statistical Institute, Kolkata, India

PhD Thesis

Title *Bayesian Modeling and Inference for Complex Dependent Non-Gaussian Data*
Advisor **Prof. Sudipto Banerjee**

Academic Appointments

- 2025–present **Postdoctoral Scholar**
Department of Statistics, University of Washington, Seattle, USA
Advisor: **Prof. Tyler H. McCormick**
- 2025–present **UW Data Science Postdoctoral Fellow**
eScience Institute, University of Washington, Seattle, USA

Awards & Achievements

- 2026 **Dean's Outstanding Student Award**, UCLA Fielding School of Public Health
[FSPH Press Release](#)
- 2026 **Inductee**, Delta Omega National Honor Society in Public Health
- 2025 **UW Data Science Postdoctoral Fellowship**, eScience Institute *USD 1,000*
- 2024 **Dissertation Year Award**, UCLA Division of Graduate Education *USD 20,000*
- 2022, 2024 **Summer Mentored Research Fellowship**, UCLA Graduate Division *USD 12,000*
- 2021 **University Fellowship**, UCLA Graduate Division *USD 14,863*
- 2016–2021 **Stipend and contingency grants**, Indian Statistical Institute
- 2015 **2nd**, Advanced Mathematical Ability Test, Calcutta Mathematical Society

Scholarly Manuscripts

* *Joint first authors*

- Published*
1. **Pan, S.**, Zhang, L., Bradley, J. R., & Banerjee, S. (2025). Bayesian inference for spatial-temporal non-Gaussian data using predictive stacking. *Bayesian Analysis*. In Press. DOI: [10.1214/25-BA1582](https://doi.org/10.1214/25-BA1582).
Selected for presentation at "Selected Papers from Bayesian Analysis" session at ISBA 2026 World Meeting, Nagoya, Japan.
 2. **Pan, S.** & Banerjee, S. (2026). Bayesian inference for spatially-temporally misaligned data using predictive stacking. *Environmetrics*, **37**(2), e70072. DOI:

[10.1002/env.70072](https://doi.org/10.1002/env.70072).

3. **Pan, S.**, Das, D., Ramachandran, G., & Banerjee, S. (2024). Bayesian hierarchical modeling and inference for mechanistic systems in industrial hygiene. *Annals of Work Exposures and Health*, **68**(8), 834–845. DOI: [10.1093/annweh/wxae061](https://doi.org/10.1093/annweh/wxae061).

Preprints

4. Yang, J. *, **Pan, S.** *, Lim, H. S., Chu, Y., Guo, Y., Agarwal, N., Babbar, V., Parikh, G. R., Chen, Y. T., Rees, C. A., Dangor, Z., Lala, S. G., Li, Z. R., Clark, S. J., Wu, Z., Datta, A., Liu, L., Rudin, C., Scarpino, S. V., Gyori, B. M., & McCormick, T. H. (2026). A Multimodal Benchmark for Evaluating Cause-of-Death Inference Using Child Health and Mortality Data. *medRxiv*. DOI: [10.64898/2026.07.13.26357980](https://doi.org/10.64898/2026.07.13.26357980)
5. Fan, Y. *, **Pan, S.** *, & McCormick, T. H. (2026). *Rashomon-Seeded Annealing for Robust Bayesian Inference in Factorial Designs*. DOI: [10.48550/arXiv.2606.02589](https://doi.org/10.48550/arXiv.2606.02589). arXiv: [2606.02589](https://arxiv.org/abs/2606.02589) [stat.ME]
6. Michel, S. K. F., **Pan, S.**, Banerjee, S., Detels, R., Giang, H. T. N., Grassman, J., & von Ehrenstein, O. S. (2026). "Province-level TCDD contamination estimates and reductions in infant birth weight in Vietnam". In: *American Public Health Association Conference* (San Antonio, Texas, United States). DOI: [10.17605/OSF.IO/PNWG7](https://doi.org/10.17605/OSF.IO/PNWG7). Submitted.
7. Michel, S. K. F., **Pan, S.**, Banerjee, S., Giang, H. T. N., Grassman, J., & von Ehrenstein, O. S. (2026). "Modeling human TCDD exposure patterns in Vietnam: Creation of a systematic database and Bayesian areal kriging". In: *International Society for Environmental Epidemiology Conference* (Munich, Bavaria, Germany). DOI: [10.17605/OSF.IO/K5ZW9](https://doi.org/10.17605/OSF.IO/K5ZW9). Accepted.
8. **Pan, S.** & Banerjee, S. (2024). *spStack: Practical Bayesian Geostatistics Using Predictive Stacking in R*. [Link](#).

Software

- CODA** AI-enabled cause of death determination assistant, *Python*, Internal Deployment, Core Algorithm Developer
- Built an automated cause-of-death classification pipeline using verbal autopsy and clinical data, benchmarking classical ML and LLMs for U5 child mortality surveillance.
 - Designed a retrieval-augmented constrained generation architecture combining causal knowledge graphs, semantic label embeddings, and clinical reasoning for ICD-11 coding.
- spStack** Bayesian geostatistics using predictive stacking, *R package*, Available on [CRAN](#).
- Fits diverse Bayesian spatial(-temporal) models for point-referenced Gaussian, Poisson, binomial, and binary data. Features a parallelizable algorithm faster than standard MCMC, built on a C++ backend with FORTRAN-optimized linear algebra.
- spStackCOS** Bayesian inference for spatial-temporal change of support models using predictive stacking, *R package*, Available on [GitHub](#).
- Implements MCMC-free Bayesian inference for spatially-temporally misaligned outcome and exposure data, featuring dedicated utilities for downstream association studies.
- modularBayes** Modular Bayesian inference for a linear model, *R package*, Available on [GitHub](#).
- A lightweight tool for Bayesian linear regression that propagates uncertainty across sequential models by using posterior distribution of predictors instead of point estimates.

Conference Presentations

- American Society of Tropical Medicine and Hygiene (ASTMH) 2026 Annual Meeting, National Harbor, Maryland, USA, November 2026. **Invited Talk.**
Using AI to improve diagnosing cause of death.
- Joint Statistical Meetings, Boston, USA, August 2026. **Invited Talk.**
Bayesian inference for spatial-temporal non-Gaussian data using predictive stacking.
- Theory and Foundations of Statistics in the Era of Big Data, Tallahassee, USA, April 2024. **Poster.**
Bayesian inference for spatial-temporal count data using predictive stacking.
- The Bayesian Young Statisticians Meeting, *j-ISBA*, Virtual Meeting, November 2023. **Short Talk.**
Predictive stacking in Bayesian hierarchical models for geospatial data.
- Joint Statistical Meetings, Toronto, Canada, August 2023. **Contributed Talk.**
Recent advances in uncertainty quantification for complex systems.
- UCLA-UCI-KAUST Meeting, Los Angeles, USA, March 2023. **Contributed Talk.**
Uncertainty quantification of dynamical systems in industrial hygiene.

Professional Experience

- 2021–2025 **Graduate Student Researcher**, *Department of Biostatistics*, UCLA
2023–2024 **Teaching Assistant**, *Department of Biostatistics*, UCLA
2021–2023 **Special Reader**, *Department of Biostatistics*, UCLA
2019–2021 **Instructor for Competitive Mathematics**, RKMV Narendrapur, India
Summer 2019 **Short-term Trainee**, National Institute of Biomedical Genomics, Kalyani, India

Teaching Experience

Education Level: G = Graduate (MS/PhD), PC = Pre-college (Grade 11/12)

- Spring 2024 250C: Multivariate Biostatistics (G), *Teaching Assistant*, UCLA
Winter 2024 250B: Linear Statistical Models (G), *Teaching Assistant*, UCLA
Fall 2023 241: Spatial Modeling and Data Analysis (G), *Teaching Assistant*, UCLA
2022–2023 255B: Advanced Probability and Statistics (G), *Special Reader*, UCLA
2021–2022 255A: Measure Theoretic Probability (G), *Special Reader*, UCLA
2019–2021 Topics in Competitive Mathematics (PC), *Instructor*, RKMV Narendrapur
Combinatorics, number theory, inequalities, polynomials, elementary real analysis etc.

Professional Memberships

(Active memberships in professional societies for research dissemination and networking.)

- International Society for Bayesian Analysis (ISBA)
- American Statistical Association (ASA)
- International Indian Statistical Association (IISA)
- Institute of Mathematical Statistics (IMS)

Referee Service

- Journal of Machine Learning Research, JMLR (1)
- Annals of Applied Statistics (3)
- Conference on Neural Information Processing Systems

Professional Service

- *Statistical Modeling Mentor*, Summer Institute in Computational Social Science (SICSS-UW), University of Washington, 2026.
- *Co-Chair*, UW Data Science Postdoctoral Fellows, eScience Institute, University of Washington, 2025–2026.
- *Volunteer*, 14th International Conference on Bayesian Nonparametrics (BNP14), Los Angeles, USA.

Technical Skills

Programming	R, C, C++ (and integration with R), Python, and Julia.
Environments	RStudio, SAS, STATA, RStudio, Jupyter, Quarto.
Infrastructure	Macintosh, Linux, Remote HPC Cluster, Slurm.
Development	Git, Continuous Integration (CI), UNIX command-line interface (CLI).
Publishing	L ^A T _E X, Markdown, MS Office Suite.
Computation	BLAS, LAPACK, Convex optimization (CVX).
Other	LEAN Proof Assistant.

Reference

Prof. Sudipto Banerjee, *PhD advisor*

Senior Associate Dean for Academic Programs
UCLA Fielding School of Public Health
Professor and Past-Chair
UCLA Department of Biostatistics
Professor of Statistics and Data Science
University of California Los Angeles
✉ sudipto@ucla.edu
🌐 <http://sudipto.bol.ucla.edu/>
☎ +1 (310) 825 5916

Last updated: July 1, 2026