Paul-Hieu V. Nguyen

January 2026

paulh_nguyen@yahoo.com

EDUCATION

University of Wisconsin–Madison, Madision, WI
Ph.D., Statistics, August 2022 - Present (Anticipated 05/27)
Reed College, Portland, OR. B.A., Mathematics, May 2022
Thesis: "Inference for Random Forests"

SKILLS

Programming: R, C++, Python, Latex, SQL. **Software**: Stan, Git
Interests: Statistical machine learning, Bayesian modeling, Uncertainty quantification, Tree ensemble methods, Simulation

PUBLICATIONS

Nguyen, P-H.V., Yee R., Deshpande, S.K. (2025). "Oblique Bayesian additive regression trees". *Transactions on Machine Learning Research.*
Wojcik O.C., Olson S.D., **Nguyen, P-H.V.**, McConville K.S., Moisen G.G. and Frescino T.S. (2022). "GREGORY: A Modified Generalized Regression Estimator Approach to Estimating Forest Attributes in the Interior Western US." *Frontiers in Forests and Global Change.*

PREPRINTS

Nguyen, P-H.V., Smoliga J.M., Lindaman B., Deshpande, S.K. (2026) "Quantifying the limits of human athletic performance: A Bayesian analysis of elite decathletes"

EXPERIENCE

Research Assistant, University of Wisconsin-Madison Aug 2022 - Present
- Developed "Oblique" extension of Bayesian Additive Regression Trees (BART) using C++ and R.
- Presented work at machine learning and statistical conferences, published paper in machine learning journal, and built R package for implementation.
- Demonstrated improvements in predictive accuracy, on average ∼5%, over current tree ensemble methods (Random Forest, XGBoost, BART) through cross-validation experiments across diverse benchmarking datasets.
- Awarded 1st Place at the CMU Sports Analytics Research Competition for development and presentation of compositional Bayesian framework for multi-output forecasting focused on uncertainty calibration and model interpretability.
- Built data pipelines integrating large-scale real estate and energy datasets to model energy usage through surrogate models with rigorous uncertainty quantification.

NSF REU Fellow, Pennsylvania State University May 2021 - Aug 2021
- Audited data privacy in large-scale Natural Language Processing dataset, utilizing statistical techniques to identify and document vulnerabilities in conversational data.

Research Assistant, Economics, Statistics, Reed College Jan 2020 - Aug 2020
- Benchmarked machine learning models (Random Forests, XGBoost) against classical econometric methods and demonstrated superior predictive power for health economic outcome variables.
- Published statistical forestry model combining satellite imagery and plot data, achieving lower-variance estimates than standard method in 71% of sampled counties; awarded 2nd in Electronic Undergraduate Statistics Research Conference.

Iowa Biostatistics Summer Fellow, University of Iowa May 2019 - Aug 2019
- Developed Bayesian spatiotemporal model to map Lyme disease incidence using health claims data to produce epidemiological risk assessments.

HONORS & AWARDS

1st place, 2025 Carnegie Mellon Sports Analytics Conference Repoducible Research Competition - Open Track, Oct 2025
Advanced Opportunity Fellowship, University of Wisconsin–Madison, Aug 2022
2nd, June 2020 Undergraduate Research Project Competition, June 2020
2020 Economics Summer Research Mintz Award, Reed College, Aug 2020
Commendation for Excellence, Reed College, Aug 2019, 2020, 2021, 2022

SERVICE Statistics Graduate Student Association Social Chair and Secretary, University of Wisconsin–Madison 2023-Present