JIE KATE HU

Phone: (206) 910-3140
khu@hsph.harvard.edu
www.katehu.com

Mailing Address
8 Story Street Suite 380
Cambridge, MA 02138

RESEARCH INTEREST

Methods: leverage auxiliary information embedded in large complex data to 1) improve estimation precision, 2) correct unmeasured confounding bias, and 3) reduce costs and bias in data collection. Corresponding methods of interest include 1) weight calibration, 2) proximal causal inference, 3) sampling and experimental designs with new sensing technologies

Theory: Z-estimation for semiparametric inference with complex study designs

Applications: climate hazards and impact studies; epidemiology; precision agriculture; environmental sensor placement

EDUCATION

PhD University of Washington, Biostatistics, 2014
Advisors: Norman Breslow and Gary Chan

MS Harvard University, Biostatistics, 2008

BS University of Hong Kong, Biochemistry, 2006
First Class Honors (Highest)
Peking University (2002-2003)
University of California, Los Angeles (2005-2006, Exchange Study)

HONORS AND AWARDS

Leadership Award

No.1 of the Most Innovative Companies in Data Science in 2021 by Fast Company
Role: Head of Data Science at Aclima, Inc; Started the Data Science team

Fellowship and Scholarship

NIH National Research Service Award Fellowship**,** 2022
~$65,000
The Jeffries Whyman Scholarship from Harvard University, 2007
~ $30, 000
Hong Kong Special Administrative Region Government Fellowship, 2003-2006
~ $45,000

Travel Awards

National Science Foundation, 2012
American Statistical Association, 2013
Graybill Conference for Modern Survey Statistics, 2013

Honors

Dean's Honor List (Every Year), 2003-2006

ACADEMIC EXPERIENCE

Postdoctoral Fellow under National Research Service Award present
Harvard T.H. Chan School of Public Health, Boston

- Mentors in methodology research: Francesca Dominici and Eric Tchetgen Tchetgen
- Developing proximal causal inference methods to adjust unmeasured and mismeasured confounding, leveraging auxiliary information from remote sensing and electronic health records
- Mentors in statistical applications: Rachel Nethery and Danielle Braun
- Investigating health impacts of wildfire, heatwave, and flood disasters
- Visualization of co-occurrence patterns of climate disasters and their impacts

PhD Dissertation 2014
Biostatistics, University of Washington, Seattle

- Advisors: Norman Breslow and Gary Chan
- Title: "A Z-estimation System for Two-phase Sampling with Applications to Additive Hazards Models and Epidemiologic Studies"
- Core study areas: case-control and case-cohort study designs; survival model development; survey methodology; semiparametric inference; empirical process theory
- Committee members: Jon Wellner, Peter Gilbert, Ying Chen, Loveday Conquest

PROFESSIONAL EXPERIENCE

Independent Consultant 2022 to present

- Advising investors on air quality sensing technologies
- Advising business leaders on building data science capabilities in start-ups

Head of Data Science/Principal Data Scientist 2019 to 2021
Aclima Inc, San Francisco

- Started, grew, and led a data science team including 5 PhD and 2 MS level researchers in statistics, optimization, and environmental sciences; the company was ranked No.1 of the most innovative data science companies by Fast Company within two years
- Led R&D on sampling designs for mapping street-by-street air quality with mobile sensing platforms, deployed in multiple metropolitan areas

- Led R&D on statistical inferential tools, machine learning, creation of hyperlocal air quality maps serving multiple regulatory agencies and environmental justice communities
- Supervised R&D on operation optimization, propagation of uncertainty, and data quality monitoring
- Forged collaborations with atmospheric scientists, engineers, and medical researchers across academia, governments, and industry
- Co-developed a capstone course project with instructors from the University of Washington --- *using hyper-local air pollution data to advance environmental justice*
- Key skills: leadership, statistics communication, data science research, big data collection, processing, and delivery

Senior Quantitative Researcher 2014 to 2019
The Climate Corporation, a subsidiary of Bayer, San Francisco

Research lead on measurement and models 2017 to 2019

- Analyzed large complex data from yield monitor, soil surveys, remote sensing imagery, environmental sensors, farm management records, weather forecast models and biogeochemical processes models
- Led cross-functional modeling efforts for location-specific and weather-adaptive nitrogen fertilizer treatment algorithms to help farmers maximize crop yield amid climate change while reducing nitrogen pollution
- Led causal inference and causal discovery research to optimize nitrogen fertilizer treatment with observational data
- Developed novel methods to integrate real-time environmental measurements, biogeochemical process models, and Bayesian hierarchical modeling results for precision agriculture management
- Secured budget to implement and evaluate the developed precision treatment algorithms in large-scale field trials over multiple states in Midwest
- Collaborated with sensor engineers to test multiple novel environmental sensing technologies on farm, conducted measurement variability studies, and developed soil measurement protocols

Research lead on experimental and sampling design 2016 to 2019

- Designed efficacy trials on farms for precision treatment algorithms evaluation
- Developed methods to 1) adapt field trials data for new hypotheses testing and 2) estimate average treatment effect with imperfect experimental data
- Designed several sampling methods for model calibration and validation, used for soil nutrients and plant tissue data collection
- Designed an adaptive sampling method for moisture and precipitation sensor placement
- Developed experimental and sampling design software TCCDesign
- Collaborated with field scientists, data engineers, sensor engineers, and farmers to implement 1-2 study designs on farm every year

- Lead authors/co-authors of 10+ internal peer-reviewed research papers and three patents

Research Assistant
Fred Hutchinson Cancer Research Center, Seattle 2008 to 2014
- Developed semi-parametric hazards models with applications to a HIV/AIDS study
- Investigated gene-gene interaction tests for Genome Wide Association Studies
- Evaluated prediction capability of biomarkers and examined biases of ROC curve

Consultant
School of Medicine, University of Washington, Seattle 2012 to 2014
- Consultation for the design and analysis of pharmaco-epidemiology studies

TEACHING EXPERIENCE

Biostatistics (Data Science Focus) Capstone II - Project Implementation 2021 Winter
- Mentor/Teacher, Aclima, Inc & University of Washington, Seattle
- Provided guidance on data visualization, data science best practice, and data science communication

Biostatistics (Data Science Focus) Capstone I - Project Planning 2020 Fall
- Mentor/Teacher, Aclima, Inc & University of Washington, Seattle
- Co-developed the course curriculum, created data sharing agreement, provided data
- Provided guidance on problem formulation, analysis of large complex spatiotemporal data, data quality management

Medical Biometry II
- Teaching Assistant at the University of Washington, Seattle 2012 Spring

Principles of Biostatistics
- Teaching Assistant at Harvard School of Public Health 2007 Fall

MENTORING EXPERIENCE (ENVIRONMENTAL DATA SCIENCE AND STATISTICS)

Doctoral Students
Dafne Zorzetto, Statistics @ University of Padova, Italy
"Bayesian nonparametric methods to adjust for unmeasured confounding" present

Sarika Aggarwal, Biostatistics @ Harvard University
"A national study of flood and its health impacts" present

Zeshi Zheng, Civil Engineering @ UC, Berkeley
"Machine learning for soil moisture sensor placement" 2017 summer, 2018 spring
Currently data science manager at C3 AI

Master Student

Cathy Chen, Biostatistics @ University of Washington, Seattle
"Health score based on hyperlocal air quality maps" 2020 summer, 2021 spring
Currently data scientist at Abbot

Undergrad Student Projects
Yibin Xiong, Applied and Computational Mathematics @ USC 2022 summer
"Performance of kernel methods for removing unmeasured confounding bias"

Josh Yamamoto, Statistics and Mathematics @ Reed College
Julian Schmitte, Applied Mathematics @ Harvard College
"Improving forest inventory estimates after wildfire" 2022 summer

Maxwell J.D. VanLandschoot, Economics @ Reed College
Julian Schmitte, Applied Mathematics @ Harvard College
"Visualizing climate change's impact on US forests" 2022 summer

GRANT APPLICATION

Google's AI for Social Good 2022
Funding for academic-NGO collaborative research
Accepted to collaborate with NGOs and proceeded to 2nd round joint proposal competition

Harvard Chan-NIEHS for Environmental Health Pilot Project 2022
"Development and Dissemination of Negative Controls to Adjust for Unmeasured Confounding: Untangle the Effects of Wildfire Smoke and Other Air Pollutants on Health"
Rejected

PATENTS (STATISTICAL AND MACHINE LEARNING METHODS)

- Hu, J. System and Method for Modular Design of Statistical Models (U. S. Application Serial No. 18/051,403)
- Hu, J, Jerkins, J, Goebel, N. Routing Method for Mobile Monitoring Platforms. (U. S. Application Serial No.17/332789)
- Hu, J & Carrion C. (2022) Using Causal Learning Algorithms to Assist in Agricultural Management Decisions. (U.S. Patent No. #11,406,053)
- Hu, J & Ladoni, M. (2021) Location Selection for Treatment Sampling ---A field Study Design Tool to Optimize Treatment Assignment and Soil Sampling Locations for Model Calibration. (U.S. Patent No. #10,963,606)
- Hu, J. (2021). Location Selection for Model Assessment. (U.S. Patent No. #10, 990,716)

COMPLETED MANUSCRIPTS

- Jie Kate Hu, Eric Tchetgen Tchetgen, Francesca Dominici, "Leveraging Auxiliary Information to Adjust for Unmeasured Confounding in Time Series Study Designs" (in revision for Nature Review Method Primer)

- Zixu Zhao, Melissa Lunden, Jie Kate Hu, Brian Lafranchi, Yutong Liang, Caleb Arata, Erin Katz, Allen H. Goldstein, Haofei Zhang, "Air Pollution Mapping and Machine Learning Reveal Key Factors for Distinct Community-level Exposures in San Francisco, California" (submitted to Environmental Science & Technology)

- Keith R. Spangler, Quinn H. Adams, Jie Kate Hu, Danielle Braun, Kate R. Weinberger, Francesca Dominici, and Gregory A. Wellenius, "Does Choice of Outdoor Heat Metric Affect Heat-Related Epidemiologic Analyses in the US Medicare Population?" (submitted to Environmental Epidemiology)

- Julian Schmitt, Josh Yamamoto, Kelly McConville, Grayson White, George Gaines, Jie Kate Hu "Zero Inflation in Small Area Estimation Models: Improving Forest Inventory Estimates" (to be submitted)

- Jie Kate Hu, Gary Chan "Weights Determination in Generalized Case-Cohort Study" (to be submitted)

- Jie Kate Hu, Moslem Ladoni "A Field Study Design Tool for Agricultural and Environmental System Model Calibration and Testing" (in submission)

- Jie Kate Hu, "A Z-estimation System: A Modular Approach to Model Development" (in submission)

PUBLICATIONS

Book Chapter

- Norman Breslow and Jie Kate Hu, "Survival Analysis of Case-Control Data: A Sample Survey Approach", Handbook of Statistical Methods for Case-Control Studies, Chapman and Hall/CRC, (2018)

Peer-Reviewed Journal Publications

Theory

- Norman E. Breslow, Jie Hu, Jon A. Wellner, "Z-estimation and Stratified Samples: Application to Survival Models", Lifetime Data Analysis 21, 493-516 (2015).

- Jie Kate Hu, Xianlong Wang, Pei Wang, "Testing Gene-gene Interactions in Genome Wide Association Studies", Genetic Epidemiology 38, 123-134 (2014)
    - Pointed out the flaws in some comparison studies of gene-gene interaction testing methods; studied definitions underlying different testing methods and provide conditions of when these tests are comparable

Methodology

- Jie Hu, Norman E. Breslow, Gary Chan, Couper David, "Estimating the Hazard Difference from Case-Cohort Studies", European Journal of Epidemiology 36(11), 1129-1142, (2021).

Applications

- Afzali Anita, Christopher J. Park, Kehao Zhu, Jie Kate Hu, Prachi Sharma, Mika N. Sinanan, and Scott D. Lee, "Preoperative Use of Methotrexate and the Risk of Early Postoperative Complications in Patients with Inflammatory Bowel Disease", Inflammatory Bowel Diseases 22(8), 1887-95 (2016).

- Afzali, Anita, Chelle L. Wheat, Jie Kate Hu, John E. Olerud, and Scott D. Lee, "The Association of Poriasiform Rash with anti-Tumor Necrosis Factor (anti-TNF) Therapy in Inflammatory Bowel Disease: A Single Academic Center Case Series", Journal of Crohn's and Colitis 8(6), 480-488 (2014).

Internal Peer-Reviewed Technical Reports, The Climate Corporation

Methodology and Applied Statistics in Precision Agriculture

- Kate Hu, Camila Casquilho, Megan Chen, Combining Measurements and Models for Nitrogen Management. Technical Report, 2018

- Kate Hu, Adjust Measurement-based Nitrogen Management Decisions using Biogeochemical Process Models. Technical Report, 2018

- Camila Casquilho, Kate Hu, Megan Chen, A Bayesian Hierarchical Model for Critical Nitrate Estimation. Technical Report, 2018

- Carlos Carrion, Kate Hu, Andrew McGowan, Megan Chen, Counterfactual Estimation of Yield Response as a Function of Soil Nutrients. Technical Report, 2018

- Camila Casquilho, Kate Hu, Spatial Variability of Pre-sidedress Nitrate. Technical Report, 2018

- Zeshi Zheng, Kate Hu, Mike Malone, Nicholas Vogel, A Time-Series Clustering Approach for Soil Moisture Probes Placement. Technical Report, 2017

- Kate Hu, A Model-Assisted Probability Sampling Design for Representative Locations. Technical Report, 2017

- Kate Hu, A Sampling Design for Model Assessment. Technical Report, 2016

- Kate Hu, Moslem Ladoni, A Sampling and Treatment Placement Tool for the Climate Corporations Nitrogen Trials. Technical Report, 2016

- Jing Cao, Kate Hu, Agronomist Survey Design and Analysis, Technical Report, 2015

- Kate Hu, Evaluation of the 2015 Climatology Field Experiments. Technical Report, 2015

- Kate Hu, Variable Seeding Rate Corn Research Partner Trials Analysis, Technical Report, 2015

PUBLIC SOFTWARE AND VISUALIZATION DASHBOARD

- Jie Hu, Fit Additive Hazards Models for Survival Analysis, CRAN - Package addhazard , (2017)

- Maxwell J.D. VanLandschoot, Julian Schmitte, (mentored by Kelly McConville & Jie Kate Hu), Climate and Tree Species Distributions Change – Informing Forestry professionals, Conservationists, and Regulators the Potential Impacts of Climate Change on US Forests, https://mjdvl.shinyapps.io/NCASI_APP/, (2022)

SELECTED TALKS

- Use Negative Controls to Adjust for Unmeasured Confounding with Time Series Studies in Environmental Epidemiology, 35th New England Statistics Symposium, Storrs, CT, 05/22 (Invited)

- Air Quality Assessment from Mobile Sensing Platform, Electrical and Computer Engineering, Stony Brook University, New York, 4/21 (Invited)

- Representative Sampling Method for Air Quality Monitoring with Mobile Sensing Platform, Air & Waste Management Association 113th Annual Conference, San Francisco, CA, 06/20

- Application of Causal Bayesian Networks to Environmental Data, Atlantic Causal Inference Conference, Pittsburgh, PA, 05/18

- Hypothesis Formulation, Experimental Design, and Analysis of Precision Agriculture Trials, Women in Statistics and Data Science, La Jolla, CA, 10/17 (Invited)

- Stories of Success, Lessons Learned, and Advice for Productive and Enjoyable Collaborations, Chair of the Panel Talk, Women in Statistics and Data Science, La Jolla, CA, 10/17 (Invited)

- Using the Additive Hazards Model with Two-Phase Sampling in Atherosclerosis Risk in Community Study, John Hopkins University, Baltimore, MD, 9/16 (Invited)

- Z-estimation for General Two-phase Sampling Problems, National Cancer Institute, Bethesda, MD, 6/14 (Invited)

- Application of Z-estimation Theory to Calibrated Estimators for Semi-parametric Models with Two-phase Stratified Sampling, Graybill Conference on Modern Survey Statistics, Fort Collins, CO, 6/13

- Parametric and Semi-parametric Analysis of Mean Residual Life Acceleration, 8th World Congress on Probability and Statistics, Istanbul, Turkey, 6/12

COMMUNITY SERVICE

- Co-chair: University of Washington Statistics in the Community (2012-2014)

- Member: Student-Faculty Communications Committee, Department of Biostatistics at the University of Washington (2012-2013)

- Member: Computer-Policy Committee, Department of Biostatistics at the University of Washington (2008-2009)

COMPUTER SKILLS

Programming: Python, R, Spark, SAS, STATA, UNIX, Scala, C++, Mathematica, Git, LaTeX, Sphinx, PlantUML, BigQuery SQL,

Platforms: Google Cloud, Google Earth, AWS, Docker, Juypter, Anaconda

RECREATIONAL ACTIVITIES

- Boston Fencing Club, Boston, 2022-now
- Halberstadt Fencers' Club, San Francisco, 2017-2021
- The Mountaineers Club, Seattle, 2010-2014
- Intramural Sports, University of Washington, Seattle

  Champion in Badminton Women's Doubles Tournament (2010, 2012)

  Champion in Co-Rec Softball League Tournament (2012)