

Xiao Hui Tai

✉ xtai@ucdavis.edu ☎ 412-557-2835 🌐 [xhtai](https://github.com/xhtai) 🌐 [xhtai.github.io](https://github.com/xhtai)

Updated: March 28, 2024

Academic Appointments

University of California, Davis, Department of Statistics

Assistant Professor

July 2022 - Present

University of California, Berkeley, School of Information

Postdoctoral Researcher

Jan 2020 - June 2022

Education

Carnegie Mellon University

Ph.D. in Statistics

Aug 2015 - May 2019

Thesis: *Forensic Data Matching Problems*

Master of Statistical Practice

Received May 2012

Singapore Management University

B.Sc. in Economics, *summa cum laude*

Received May 2011

Publications (Peer-reviewed)

Yury Elena Garcia Puerta, Miryam Elizabeth Villa-Perez, Kuang Li, **Xiao Hui Tai**, Luis A. Trejo, Maria L. Daza-Torres, J. Cricelio Montesinos-Lopez and Miriam Nuno. Wildfires and Social Media Discourse: Exploring Mental Health and Emotional Well-Being Through Twitter. *Frontiers in Public Health*. 2024 (in press).

Arogya Koirala, Suraj R. Nair and **Xiao Hui Tai**. [Mapping Opium Poppy Cultivation: Socio-economic Insights from Satellite Imagery](#). *ACM Journal on Computing and Sustainable Societies*. 2024.

Early versions: 4th ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '21) (poster track); 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), Workshop on Humanitarian Mapping

Xiao Hui Tai, Shikhar Mehra and Joshua E. Blumenstock. [Mobile Phone Data Reveal the Effects of Violence on Internal Displacement](#). *Nature Human Behavior*. 2022.

Cornelia Ilin*, Sébastien Annan-Phan*, **Xiao Hui Tai***, Shikhar Mehra, Solomon Hsiang and Joshua E. Blumenstock. [Public Mobility Data Enables COVID-19 Forecasting and Management at Local and Global Scales](#). *Nature Scientific Reports*. 2021. (* indicates equal contribution)

Xiao Hui Tai and Kayla Frisoli. [Benchmarking Minimax Linkage in Hierarchical Clustering](#). *Data Analysis and Rationality in a Complex World*. Springer International Publishing, 2021.

Xiao Hui Tai, Kyle Soska and Nicolas Christin. [Adversarial Matching of Dark Net Market Vendor Accounts](#). *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*. 2019.

Xiao Hui Tai. [Record Linkage and Matching Problems in Forensics](#). *IEEE 18th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018.

Xiao Hui Tai and William F. Eddy. [A Fully Automatic Method for Comparing Cartridge Case Images](#). *Journal of Forensic Sciences*. 2018.

Book Chapters and Magazine Articles

Susan VanderPlas, Alicia Carriquiry, Heike Hofmann, James Hamby and **Xiao Hui Tai**. [An Introduction to Firearms Examination for Researchers in Statistics](#). *Handbook of Forensic Statistics*. 2021.

Sam Tyner, Soyoung Park, Ganesh Krishnan, Karen Pan, Eric Hare, Amanda Luby, **Xiao Hui Tai**, Heike Hofmann, and Guillermo Basulto-Elias. 2019. [OpenForSciR: Open Forensic Science in R](#). 2019.

Alicia Carriquiry, Heike Hofmann, **Xiao Hui Tai** and Susan VanderPlas. [Machine Learning in Forensic Applications](#). *Significance*. 2019.

Pre-prints

Xiao Hui Tai, Suraj R. Nair, Shikhar Mehra and Joshua E. Blumenstock. Satellite and Mobile Phone Data Reveal How Violence Affects Seasonal Migration in Afghanistan. *In submission*.

Pablo Busch, Paulo Rocha, Kyung Jin Lee, Luis Abdón Cifuentes and **Xiao Hui Tai**. Acute exposure to fine particulate pollution and elderly mortality in Chile. *In submission*.

Xiao Hui Tai and William F. Eddy. [Automatically Matching Topographical Measurements of Cartridge Cases Using a Record Linkage Framework](#). *arXiv:2003.00060*.

Research Funding

UC Davis Clinical and Translational Science Center, Co-Investigator (2023). Title: Mental Health and Well-Being Effects of Wildfire Smoke: A Translational Research Framework for Action (PI: Miriam Nuño). Budget: \$50,000.

Professional Experience

Quantitative Modeler, JPMorgan Chase & Co. June 2019 - Jan 2020
Develop and maintain models to forecast movements between balance tiers in checking accounts, for use in stress testing.

Statistician, Ministry of Trade and Industry, Singapore Oct 2012 - August 2015
Maintained a database of demographic information and collaborated with economists and social scientists on policy-related statistical analysis.

Awards and Honors

- 2019 Honorable Mention for Cluster Benchmark Challenge at the 16th Conference of the International Federation of Classification Societies
- 2019 Center for Statistics and Applications in Forensic Evidence Poster Award
- 2017 American Statistical Association Travel Award
- 2011 Monetary Authority of Singapore Academic Excellence Award
- 2011 School Salutatorian Award

Teaching and Advising

Teaching: STA 35A Statistical Data Science 1 (Fall 2022, Fall 2023), STA 160 Practice in Statistical Data Science (Fall 2023), STA 141C Big Data and High Performance Statistical Computing (Spring 2024), STA 250 Data Science for International Development (Winter 2023), STA 290 Statistics Seminar (Winter 2023, Fall 2023)

Advising:

- Ph.D. advisor: Oscar Rivera (ongoing)
- Ph.D. Committee Member: Abdul-Hakeem Omotayo (ongoing)
- Masters: Arogya Koirala (Stanford)
- Undergraduates: Dan Xiang, Tej Sathe, Neda Shahiar, Musa Jalis

Teaching assistant: Introduction to Statistical Inference (Spring 2012), Modern Regression (Fall 2018), Undergraduate Capstone (Fall 2018)

Guest lecturer: Modern Regression (Fall 2018), Big Data and Development (Spring 2021)

Selected Presentations

Armed Conflict and Girls' Education in Africa

- (Accepted) South-East Exchange of Development Studies (SEEDS) 4th Annual Conference, Atlanta, GA. 04/2024.
- (Accepted) 21st Midwest International Economic Development Conference, Chicago, IL. 04/2024.
- (Accepted) 19th Annual Workshop of the Households in Conflict Network, Brussels, Belgium. 11/2023.

Civil Conflict Affects Seasonal Labor Flows

- (Invited) Empirical Studies of Conflict Workshop on Migration and Forced Displacement: Research to Inform Policy, Washington, D.C. 10/2023.
- (Invited) Stanford University Statistics Seminar, Stanford, CA. 07/2023.
- (Accepted) Asia School of Business-UNHCR Research Workshop on Refugee Studies and Forced Displacement, Kuala Lumpur, Malaysia. 06/2023.
- (Accepted) Empirical Studies of Conflict Latin American Conference: Bogotá, Colombia. 05/2023.
- (Invited) UC Santa Cruz Statistics Seminar, Santa Cruz, CA. 05/2023.
- (Invited) UC Davis Statistics Department Seminar, Davis, CA. 10/2022.
- (Accepted) 22nd Meeting of New Researchers in Statistics and Probability, Fairfax, VA. 08/2022.
- (Accepted) World Bank-Center for Effective Global Action (CEGA) Measuring Development Conference 2022, Washington, D.C. 05/2022.

The Right to Privacy: Consumer Genetics, Data Brokers, Facial Recognition Technology, Criminal Justice, and You. (Invited panel) Science Ethics and Policy Symposium 2022, Berkeley, CA. 04/2022.

Mobile Phone Data Reveal the Effects of Violence on Internal Displacement.

- (Invited) Lancaster University Computation in AI and Social Science Hub, virtual. 12/2023.
- (Invited) Yonsei University Department of Statistics and Data Science, Seoul, Korea. 03/2023.
- (Invited) UMass Amherst (Statistics and Data Science Seminar Series), virtual. 01/2022.
- (Accepted) World Bank-UNHCR 2nd Research Conference on Forced Displacement, virtual. 01/2022.
- (Accepted) American Statistical Association Joint Statistical Meetings, virtual. 08/2020.

Mapping Opium Poppy Cultivation in Afghanistan Using Satellite Imagery.

- (Accepted) 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Workshop on Humanitarian Mapping, virtual. 08/2021.
- (Accepted) COMPASS '21: ACM SIGCAS Conference on Computing and Sustainable Societies (poster track), virtual. 06/2021.

Benchmarking Minimax Linkage in Hierarchical Clustering. (Invited talk) 16th Conference of the International Federation of Classification Societies (IFCS), Thessaloniki, Greece. 08/2019.

Record Linkage and Matching Problems in Forensics. IEEE 18th International Conference on Data Mining Workshops (ICDMW), Singapore. 11/2018.

Matching Seller Accounts on Online Anonymous Marketplaces.

- (Accepted) 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Anchorage, Alaska. 08/2019.
- (Accepted) American Statistical Association Joint Statistical Meetings, Vancouver, BC, Canada. 08/2018.
- (Invited) Cyber Analytics Workshop at CyLab Partners Conference 2017, Pittsburgh, PA. 09/2017.
- (Accepted) American Statistical Association Women in Statistics and Data Science Conference 2017, La Jolla, CA. 10/2017.
- (Invited) Memorial Invited Special Session: Remembering Stephen E. Fienberg at the International Conference on Forensic Inference and Statistics, Minneapolis, MN. 09/2017.

Automatic Comparison of Cartridge Cases.

- (Invited) American Bar Association Ninth Annual Prescription for Criminal Justice Forensics Program, New York, NY. 05/2018.
- (Invited) Center for Statistics and Applications in Forensic Evidence (CSAFE) Center Wide Webinar. 01/2018.
- (Invited) Firearms Examiner Workshop, Ames, IA. 12/2017.
- (Invited) Workshop on Role of Data, Databases and Expert Knowledge in Forensic Inference, Pittsburgh, PA. 09/2017.
- (Accepted) American Statistical Association Joint Statistical Meetings, Baltimore, MD. 08/2017.
- (Accepted) Association of Firearms and Toolmarks Examiners Conference, Denver, CO. 05/2017.
- (Invited) ForensicsNIST, Gaithersburg, MD. 11/2016.

[Assessing Contagious Spread of Competing Hashtags on Twitter](#). Poster at the Public Exhibition of LARC Posters and Demos, Singapore. 10/2012.

Service

Reviewer

- 2023 Journal of Computational and Graphical Statistics, Nature Human Behavior, ACM COMPASS
- 2022 Proceedings of the National Academy of Sciences
- 2021 PloS one, Journal of Forensic Sciences
- 2020 C3.ai Digital Transformation Institute
- 2018 Journal of Forensic Sciences

Mentorship and Teaching

- 2021 Research Mentor, De Anza High School
- 2020 SAT Math Instructor, Berkeley
- 2018-19 PhD Mentor, CMU Department of Statistics
- 2017 Panel on Applying to Graduate School, CMU

Others

- 2022 Volunteer, American Causal Inference Conference
- 2019 Consultant at Ministry of Trade and Industry, Singapore
- 2018-19 Volunteer, Pittsburgh Women in Statistics Conference
- 2018 Volunteer for Statistics Without Borders

Skills

Programming languages: R, SQL, SAS (proficient), Matlab (intermediate), Python, Stata (some experience)

Operating systems: OS X, Linux