

## Research Interests

Statistical machine learning and inference; developing new statistical methods for complex-data settings; using convex optimization to develop statistical methods with theoretical guarantees; structured sparsity for high-dimensional settings; developing software tools for better, more reproducible research.

## Education

*Ph.D. in Statistics*, 2012  
Stanford University (advisor: Robert Tibshirani)

*M.S. in Statistics*, 2007  
Stanford University

*B.S. in Physics “with Distinction” and Departmental Honors*, 2006  
Stanford University (secondary major in Mathematics)

## Academic Positions

*Professor*, Department of Data Sciences and Operations, University of Southern California (November 2023–present)

*Dean’s Associate Professor in Business Administration*, Department of Data Sciences and Operations, University of Southern California (August 2021–May 2024)

*Associate Professor (with tenure)*, Department of Data Sciences and Operations, University of Southern California (February 2019–November 2023)

*Assistant Professor*, Department of Data Sciences and Operations, University of Southern California (June 2017–February 2019)

*Assistant Professor*, Department of Biological Statistics and Computational Biology, Cornell University (August 2012–June 2017)

Joint appointment: Department of Statistical Science

## Selected Awards, Honors, & Grants

*Dean’s Award for Research Impact*, Marshall School of Business, University of Southern California (May 2025)

*Fellow*, American Statistical Association (April 2025)

*Fellow*, Institute of Mathematical Statistics (May 2024)

*Award in Innovation in Statistical Programming and Analytics* by the American Statistical Association’s Statistical Programmers and Analysts Section (July 2023)

Journal of the American Statistical Association Reproducibility Award (for paper “Selective Inference for Hierarchical Clustering” JASA 2022) (May 18, 2023)

Simons Foundation Collaboration on Ocean Processes and Ecology: Principal Investigator for 2-year grant. Title: “SCOPE Synthesis: Statistical Learning Along Oceanographic Transects” (September 1, 2025 – August 31, 2027)

Simons Foundation Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems: Principal Investigator for 5-year grant. Title: “CBIOMES: Statistical Learning for Marine Ecosystems” (August 1, 2022 – July 31, 2027)

*Dean's Associate Professorship in Business Administration*, Marshall School of Business, University of Southern California (June 2021)

NIH R01 grant: Principal Investigator (multiple-PI with Daniela Witten) for 4-year grant. Title: "A Modeling Framework for Multi-View Data, with Applications to the Pioneer 100 Study and Protein Interaction Networks." (August 1, 2017 – June 30, 2021)

Simons Foundation Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems: Principal Investigator for 5-year grant with 1-year supplement. Title: "CBIOMES: Statistical Time Series Analysis and Network Inference" (August 1, 2017 – July 31, 2023)

NSF CAREER Award: Principal Investigator for 5-year National Science Foundation grant. Title: "Flexible Parsimonious Models for Complex Data" (July 1, 2017–June 30, 2022)

NSF DMS grant: Principal Investigator for 3-year National Science Foundation grant. Title: "High-Dimensional Covariance Estimation via Convex Optimization." (August 15, 2014–July 31, 2017)

*Yahoo! Key Scientific Challenges Program* Award in Machine Learning and Statistics (Spring 2011)

Lieberman Fellowship (Spring 2010)

Stanford Statistics *Departmental* TA Award (Spring 2010)

Stanford Statistics TA Award (Spring 2009)

Urbanek Family Stanford Graduate Fellowship for 3 years (Spring 2007)

Inducted into Phi Beta Kappa (Spring 2006)

President's Scholar, Stanford University (Spring 2002)

## Preprints

Di Wu, Jacob Bien, and Snigdha Panigrahi (2025) **Hierarchical Clustering with Confidence**. (<https://www.arxiv.org/abs/2512.06522>)

Farhad de Sousa, Francois Ribalet, and Jacob Bien (2025) **Automated Gating for Flow Cytometry Data Using a Kernel-Smoothed EM Algorithm**. (<https://arxiv.org/abs/2510.06051>)

Yiling Huang, Snigdha Panigrahi, Guo Yu, and Jacob Bien (2025) **Reluctant Interaction Inference after Additive Modeling**. (<https://arxiv.org/abs/2506.01219>)

Xiaozhu Zhang, Jacob Bien, and Armeen Taeb (2025) **Quantifying Uncertainty and Stability Among Highly Correlated Predictors: A Subspace Perspective**. (<https://arxiv.org/abs/2505.06760>)

Sangwon Hyun, Tim Coleman, Francois Ribalet, and Jacob Bien (2025) **Trend Filtered Mixture of Experts for Automated Gating of High-Frequency Flow Cytometry Data**. (<https://arxiv.org/abs/2504.12287>)

Oh-Ran Kwon, Gourab Mukherjee, and Jacob Bien (2024) **Semi-Supervised Learning of Noisy Mixture of Experts Models**. (<https://arxiv.org/abs/2410.09039>)

Gregory Faletto and Jacob Bien (2022) **Cluster Stability Selection**. (<https://arxiv.org/abs/2201.00494>)

Guo Yu, Jacob Bien, and Ryan Tibshirani (2019) **Reluctant Interaction Modeling**. (<https://arxiv.org/abs/1907.08414>)

Jacob Bien (2016) **Simulator: An Engine to Streamline Simulations**. (<http://arxiv.org/abs/1607.00021>)

## Publications

Ameer Dharamshi, Anna Neufeld, Lucy Gao, Daniela Witten, and Jacob Bien (2025) **Thinning a Wishart Random Matrix**. *Biometrika*.

Ronan Perry, Snigdha Panigrahi, Jacob Bien, and Daniela Witten (2025) **Inference on the Proportion of Variance Explained in Principal Component Analysis**. *Journal of the American Statistical Association*.

Ameer Dharamshi, Anna Neufeld, Lucy Gao, Jacob Bien, and Daniela Witten (2025) **Decomposing Gaussians with Unknown Covariance**. *Biometrika*.

Anna Neufeld, Ameer Dharamshi, Lucy L Gao, Daniela Witten, and Jacob Bien (2025) **Discussion of “Data Fission: Splitting a Single Data Point”**. *Journal of the American Statistical Association*. **120**(549). 151–157.

Jacob Bien and Gourab Mukherjee (2025) **Generative AI for Data Science 101: Coding Without Learning to Code**. *Journal of Statistics and Data Science Education*. **33**(2). 129–142.

Adel Javanmard, Simeng Shao, and Jacob Bien (2025) **Prediction Sets for High-Dimensional Mixture of Experts Models**. *Journal of the Royal Statistical Society, Series B*. **87**(3). 850—871.

Ameer Dharamshi, Anna Neufeld, Keshav Motwani, Lucy L Gao, Daniela Witten, and Jacob Bien (2025) **Generalized Data Thinning using Sufficient Statistics**. *Journal of the American Statistical Association*. **120**(549). 511–523.

**\*\*Winner of 2024 ASA Statistical Learning and Data Science Section Student Paper Award\*\***

Arkajyoti Saha, Daniela Witten, and Jacob Bien (2025) **Inferring Independent Sets of Gaussian Variables after Thresholding Correlations**. *Journal of the American Statistical Association*. **120**(549). 370–381.

Lucy L Gao, Jacob Bien, and Daniela Witten (2024) **Selective Inference for Hierarchical Clustering**. *Journal of the American Statistical Association*. **119**(545). 332–342.

**\*\*Winner of 2022 JASA Reproducibility Award\*\***

Simeng Shao, Jacob Bien, and Adel Javanmard (2024) **Controlling the False Split Rate in Tree-Based Aggregation**. *Journal of the American Statistical Association*.

Gregory Faletto and Jacob Bien (2023) **Predicting Rare Events by Shrinking Towards Proportional Odds**. *International Conference on Machine Learning 2023*.

Sangwon Hyun, Mattias Rolf Cape, Francois Ribalet, and Jacob Bien (2023) **Modeling Cell Populations Measured by Flow Cytometry with Covariates using Sparse Mixture of Regressions**. *Annals of Applied Statistics*. **17**(1). 357–377.

Andee Kaplan and Jacob Bien (2023) **Interactive Exploration of Large Dendrograms with Prototypes**. *The American Statistician*. **77**(2). 201–211.

Bror Fredrik Jönsson, Christopher Follett, Jacob Bien, Stephanie Dutkiewicz, Sangwon Hyun, Gemma Kulk, Gael Forget, Christian Müller, Marie-Fanny Racault, Christopher Nigel Hill, Thomas Jackson, and Shubha Sathyendranath (2023) **Using Probability Density Functions to Evaluate Models (PDFEM, V1. 0) to Compare a Biogeochemical Model with Satellite Derived Chlorophyll**. *Geoscientific Model Development*. **16**(16). 4639–4657.

Ryan Reynolds, Sangwon Hyun, Benjamin Tully, Jacob Bien, and Naomi M Levine (2023) **Identification of Microbial Metabolic Functional Guilds from Large Genomic Datasets.** *Frontiers in Microbiology*. **14**.

Ines Wilms, Sumanta Basu, Jacob Bien, and David S Matteson (2023) **Sparse Identification and Estimation of Large-Scale Vector Autoregressive Moving Averages.** *Journal of the American Statistical Association*. **118**(541). 571–582.

Ines Wilms and Jacob Bien (2022) **Tree-Based Node Aggregation in Sparse Graphical Models.** *The Journal of Machine Learning Research*. **23**(1). 11078–11113.

Evan L. Ray, Logan C. Brooks, Jacob Bien, Matthew Biggerstaff, Nikos I. Bosse, Johannes Bracher, Estee Y. Cramer, Sebastian Funk, Aaron Gerding, Michael A. Johansson, Aaron Rumack, Yijin Wang, Martha Zorn, Ryan J. Tibshirani, and Nicholas G. Reich (2022) **Comparing Trained and Untrained Probabilistic Ensemble Forecasts of COVID-19 Cases and Deaths in the United States.** *International Journal of Forecasting*.

Sangwon Hyun, Aditya Mishra, Christopher L. Follett, Bror Jonsson, Gemma Kulk, Gael Forget, Marie-Fanny Racault, Thomas Jackson, Stephanie Dutkiewicz, Christian L. Müller, and Jacob Bien (2022) **Ocean Mover's Distance: Using Optimal Transport for Analyzing Oceanographic Data.** *Proceedings of the Royal Society A*. **478**(2262).

Estee Y. Cramer, ..., Jacob Bien, ..., and Nicholas G. Reich (2022) **Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the US.** *Proceedings of the National Academy of Sciences*. **119**(15).

Lucy L. Gao, Daniela Witten, and Jacob Bien (2022) **Testing for Association in Multiview Network Data.** *Biometrics*. **78**(3). 1018–1030.

**\*\*Winner of 2020 ASA Statistical Learning and Data Science Section Student Paper Award\*\***

Guo Yu, Daniela Witten, and Jacob Bien (2022) **Controlling Costs: Feature Selection on a Budget.** *Stat*. **11**(1). e427.

Daniel J. McDonald, Jacob Bien, Alden Green, Addison J. Hu, Nat DeFries, Sangwon Hyun, Natalia L. Oliveira, James Sharpnack, Jingjing Tang, Robert Tibshirani, Valérie Ventura, Larry Wasserman, and Ryan J. Tibshirani (2021) **Can Auxiliary Indicators Improve COVID-19 Forecasting and Hotspot Prediction?** *Proceedings of the National Academy of Sciences*. **118**(51).

Alex Reinhart, ..., Jacob Bien, ..., Roni Rosenfeld, and Ryan Tibshirani (2021) **An Open Repository of Real-Time COVID-19 Indicators.** *Proceedings of the National Academy of Sciences*. **118**(51).

Jacob Bien, Xiaohan Yan, Leo Simpson, and Christian Müller (2021) **Tree-Aggregated Predictive Modeling of Microbiome Data.** *Scientific Reports*. **11**(14505).

Xiaohan Yan and Jacob Bien (2021) **Rare Feature Selection in High Dimensions.** *Journal of the American Statistical Association*. **116**(534). 887–900.

William B Nicholson, Ines Wilms, Jacob Bien, and David S Matteson (2020) **High Dimensional Forecasting via Interpretable Vector Autoregression.** *The Journal of Machine Learning Research*. **21**(1). 6690–6741.

Shuxiao Chen and Jacob Bien (2020) **Valid Inference Corrected for Outlier Removal.** *Journal of Computational and Graphical Statistics*. **29**(2). 323–334.

**\*\*Winner of JSM 2018 Statistical Learning and Data Science Student Paper Award\*\***

Lucy L Gao, Jacob Bien, and Daniela Witten (2020) **Are Clusterings of Multiple Data Views Independent?**. *Biostatistics*. **21**(4). 692–708.

**\*\*Winner of JSM 2019 Biometrics Section Travel Award\*\***

Guo Yu, Jacob Bien, and Daniela Witten (2019) **Discussion of “Covariate-Assisted Ranking and Screening for Large-Scale Two-Sample Inference”**. *Journal of the Royal Statistical Society, Series B*.

Guo Yu and Jacob Bien (2019) **Estimating the Error Variance in a High-Dimensional Linear Model**. *Biometrika*. **106**(3). 533–546.

Jacob Bien (2019) **Graph-Guided Banding of the Covariance Matrix**. *Journal of the American Statistical Association*. **114**(526). 782–792.

James Li, Jacob Bien, and Martin T Wells (2018) **rTensor: An R Package for Multidimensional Array (Tensor) Unfolding, Multiplication, and Decomposition**. *Journal of Statistical Software*. **87**. 1–31.

Jacob Bien, Irina Gaynanova, Johannes Lederer, and Christian L Müller (2019) **Prediction Error Bounds for Linear Regression with the TREX**. *Test*. **28**. 451–474.

Jacob Bien, Irina Gaynanova, Johannes Lederer, and Christian L Müller (2018) **Non-Convex Global Minimization and False Discovery Rate Control for the TREX**. *Journal of Computational and Graphical Statistics*. **27**(1). 23–33.

Xiaohan Yan and Jacob Bien (2017) **Hierarchical Sparse Modeling: A Choice of Two Group Lasso Formulations**. *Statistical Science*. **32**(4). 531–560.

Guo Yu and Jacob Bien (2017) **Learning Local Dependence in Ordered Data**. *Journal of Machine Learning Research*. **18**(42). 1–60.

Ines Wilms, Sumanta Basu, Jacob Bien, and David S Matteson (2017) **Interpretable Vector AutoRegressions with Exogenous Time Series**. *arXiv preprint arXiv:1711.03623*.

William B Nicholson, David S Matteson, and Jacob Bien (2017) **VARX-L: Structured Regularization for Large Vector Autoregressions with Exogenous Variables**. *International Journal of Forecasting*. **33**(3). 627–651.

Yin Lou, Jacob Bien, Rich Caruana, and Johannes Gehrke (2016) **Sparse Partially Linear Additive Models**. *Journal of Computational and Graphical Statistics*. **25**(4). 1126–1140.

Jacob Bien, Florentina Bunea, and Luo Xiao (2016) **Convex Banding of the Covariance Matrix**. *Journal of the American Statistical Association*. **111**(514). 834–845.

Jacob Bien and Daniela Witten (2016) **Penalized Estimation in Complex Models** (book chapter) in P. Bühlmann and P. Drineas and M. Kane and M. van der Laan (Eds.), *Handbook of Big Data*. Chapman and Hall/CRC Reference.

Jacob Bien, Noah Simon, and Robert Tibshirani (2015) **Convex Hierarchical Testing of Interactions**. *Annals of Applied Statistics*. **9**(1). 27–42.

Jacob Bien, Jonathan Taylor, and Robert Tibshirani (2013) **A Lasso for Hierarchical Interactions**. *Annals of Statistics*. **41**(3). 1111–1141.

Jacob Bien and Marten Wegkamp (2013) **Discussion of “Correlated Variables in Regression: Clustering and Sparse Estimation”**. *Journal of Statistical Planning and Inference*. **143**(11). 1859–1862.

Jacob Bien and Robert Tibshirani (2011) **Hierarchical Clustering with Prototypes via Minimax Linkage**. *Journal of the American Statistical Association*. **106**(495). 1075–1084.

Jacob Bien and Robert Tibshirani (2011) **Sparse Estimation of a Covariance Matrix**. *Biometrika*. **98**(4). 807–820.

Robert Tibshirani, Jacob Bien, Jerome Friedman, Trevor Hastie, Noah Simon, Jonathan Taylor, and Ryan J Tibshirani (2012) **Strong Rules for Discarding Predictors in Lasso-Type Problems**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **74**(2). 245–266.

Jacob Bien and Robert Tibshirani (2011) **Prototype Selection for Interpretable Classification**. *Annals of Applied Statistics*. **5**(4). 2403–2424.

Neema Moraveji, Daniel Russell, Jacob Bien, and David Mease (2011) **Measuring Improvement in User Search Performance Resulting from Optimal Search Tips**. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information*.

Jacob Bien, Yan Xu, and Michael Mahoney (2010) **CUR from a Sparse Optimization Viewpoint**. *Advances in Neural Information Processing Systems* 23.

## Other Writing

Ray, E., Brooks, L., Bien, J., Bracher, J., Gerding, A., Rumack, A., Biggerstaff, M., Johansson, M., Tibshirani, R(yan), Reich, N. (2021) **Challenges in training ensembles to forecast COVID-19 cases and deaths in the United States** *International Institute of Forecasters* blog post [link]

Brooks, L., Ray, E., Bien, J., Bracher, J., Rumack, A., Tibshirani, R(yan), Reich, N. (2020) **Comparing ensemble approaches for short-term probabilistic COVID-19 forecasts in the U.S.** *International Institute of Forecasters* blog post [link]

Bien, J. (2006) **Protein Motions and Mutations from a Molecular Dynamics Viewpoint**, *Undergraduate Honors Thesis*. Advisor: Sebastian Doniach

## Teaching

Two awards (see above).

Course Instructor at University of Southern California:

- DSO 699: **Advanced Data Science: Modeling, Computing, & Optimization**
  - PhD course with students in several departments (including DSO, Math, Economics, ISE)
- GSBA 524: **Data Science for Business**
  - *core course in full-time MBA program*
  - taught second-half of Fall 2022 and Fall 2023
  - integrated generative AI into Fall 2023 class and wrote paper about experience (“Generative AI for Data Science 101: Coding Without Learning to Code”)
- BUAD 312: **Statistics and Data Science for Business**
  - *developed new class; faster-paced introductory course for undergraduate business students, teaching programming (in R with tidyverse) and statistics*
  - *class described as “hugely popular” in Poets & Quants for Undergrads article “10 Undergraduate Business Schools To Watch In 2022” (Feb 26, 2022)*  
<https://poetsandquantsforundergrads.com/2022/02/26/10-undergraduate-business-schools-to-watch-in-2022/>
  - taught Spring 2020, Fall 2020, Spring 2021, Fall 2024, Fall 2025
- BUAD 310: **Applied Business Statistics**
  - *introductory class for undergraduate business students*
  - taught Fall 2017, Fall 2018

Course Instructor at Cornell University:

- BTRY 6010/ILRST 6100, **Statistical Methods I**
  - *introductory class for graduate students in fields other than statistics*
  - taught Fall 2016, Fall 2014
- BTRY 7180, **Generalized Linear Models**
  - *core course in Statistics PhD program*
  - taught Spring 2016, Fall 2013
- BTRY/STSCI 4090, **Theory of Statistics**
  - *“capstone” theoretical course of Statistics major*
  - taught Spring 2017, Spring 2016, Spring 2015, Spring 2014, Spring 2013
- BTRY 4950/7950, **Statistical Consulting**
  - *hands-on, team-taught course for graduate and undergraduate students*
  - taught Spring 2014

Course Instructor at Stanford University:

- Statistics 319, **Literature of Statistics**
  - *course for PhD students to practice reading and presenting literature*
  - taught Spring 2012

Teaching Assistant at Stanford University:

- Stat 305, **Introduction to Statistical Modeling** (PhD level, Fall 2011)
- Stat 362, **Monte Carlo Methods** (PhD level, Fall 2010)
- Stat 315A, **Modern Applied Statistics: Elements of Statistical Learning**, Head TA (PhD level, Winter 2010)
- Stat 116 **Theory of Probability** Head TA (undergraduate level, Spring 2009)
- Stat 60 **Introduction to Statistical Methods: Precalculus** (undergraduate level, Winter 2008)

## R Packages Written and Maintained

jsm2025: R package for recommending what talks to go to at JSM based on citation data (2025).

jsm2024: R package for recommending what talks to go to at JSM based on citation data (2024).

jsm2023: R package for recommending what talks to go to at JSM based on citation data (2023).

litr: R package for literate programming for writing R packages (2022).

jsm2022: R package for recommending what talks to go to at JSM based on citation data (2022).

evalcast: R package for evaluating COVID-19 forecasters (2020).

trac: R package for tree-based aggregation of compositional data (2020).

ggb: R package for graph-guided banding of the covariance matrix (2016).

simulator: R package for streamlining simulation studies (2016).

convexjulia: R package for calling Julia’s Convex.jl package from R (2015).

hierband: R package for “Convex banding of the Covariance Matrix” (2014).

protoclass: R package for “Prototype Selection for Interpretable Classification” (2013).  
 hierNet: R package for “A Lasso for Hierarchical Interactions” (2012).  
 spcov: R package for “Sparse Estimation of a Covariance Matrix,” (2012).  
 CVXfromR: R package for calling Matlab’s CVX from R (2011).  
 protoclust: R package for “Hierarchical Clustering with Prototypes via Minimax Linkage,” (2011).

## **R Packages Supervised (Written by Students/Postdocs)**

cssr: Cluster Stability Selection (with Greg Faleto)  
 hat: Hierarchical Aggregation through Testing (with Simeng Shao and Adel Javanmard)  
 flowmix: Sparse Multivariate Mixture of Regressions for Flow Cytometry Data (with Sang-won Hyun)  
 sprintr: Sparse Reluctant Interaction Modeling (with Guo Yu)  
 rare: Linear Model with Tree-Based Lasso Regularization for Rare Feature Selection (with Xiaohan Yan)  
 natural: Estimating the Error Variance in a High-Dimensional Linear Model (with Guo Yu)  
 outference: A Tool for Valid Inference Corrected for Outlier Removal in Linear Regressions (with Shuxiao Chen)  
 BigVAR: Dimension Reduction Methods for Multivariate Time Series (with Will Nicholson and David Matteson)  
 rTensor: Tools for Tensor Analysis and Decomposition (with James Li and Martin Wells)  
 hsm: A Path-Based BCD for Proximal Function of Latent Group Lasso (with Xiaohan Yan)  
 varband: Variable Banding of Large Precision Matrices (with Guo Yu)

**Presentations** Workshop (on writing R packages with `litr`) for Women in Statistics and Data Science Conference (virtual) (October 14, 2025)  
 Invited talk at EcoSta 2025 conference (virtual) (August 22, 2025)  
 Invited talk at Florida State University Statistics Department Seminar (April 25, 2025)  
 Invited talk at UC Santa Cruz Statistics Department Seminar (February 10, 2025)  
 Invited talk on generative AI in statistics and data science education for EDGE-DS Lecture Series at UC Santa Cruz (February 10, 2025)  
 Workshop (on writing R packages with `litr`) for UC Santa Cruz Statistics Department (February 10, 2025)  
 Invited talk for JSDSE/CAUSE webinar (Journal of Statistics and Data Science Education) on generative AI in statistics and data science education (February 4, 2025)  
 Invited talk at ICSDS conference in Nice, France (December 18, 2024)  
 Invited talk at SLDS conference in Newport Beach, CA (November 7, 2024)  
 Invited talk at Stanford Statistics Department Seminar (October 1, 2024)



Invited discussant in JASA Theory and Methods invited paper session at Joint Statistical Meetings in Portland, OR (August 5, 2024)

Invited presenter at invited poster session at Joint Statistical Meetings in Portland, OR (August 4, 2024)

Invited talk at Institute for Research in Statistics and its Application, University of Minnesota, conference on “New Perspectives on the Analysis of Complex Multivariate Data” (May 30, 2024)

Invited panelist for ASA's Statistics and Data Science Education Section's panel on generative AI in statistics and data science education (May 21, 2024)

Invited talk at retreat on statistics/biostatistics curriculum reform (May 17, 2024)

Invited talk at University of British Columbia Statistics Department (May 9, 2024)

Moderator for industry panel at USC Data Science and Operation's SEEDS conference (January 2024)

Invited talk at 2023 Princeton Day of Statistics (November 3, 2023)

Invited moderator for panel on “New and Evolving Roles of Business Statistics in the Big-data Era” at Joint Statistical Meetings in Toronto, Canada (August 9, 2023)

Invited talk at PDT Partners Academic Colloquium Series (June 27, 2023)

Invited talk at Memorial Sloan Kettering Cancer Center Biostatistics Seminar (June 26, 2023)

Workshop (on writing R packages with `litr`) for Ludwig Maximilian University of Munich Statistics Department (June 13, 2023)

Invited talk at Ludwig Maximilian University of Munich Statistics Department (June 12, 2023)

Invited talk at “Workshop on Challenges for Statistics in the Era of Data Science” in Hannover, Germany (June 7–9, 2023). Talk title: “The Role of *Human* Statisticians in the Future of Data Science.”

Invited talk at UC Irvine Statistics Seminar (May 25, 2023)

Invited talk at UC Riverside Statistics Seminar (May 9, 2023)

Workshop (on writing R packages with `litr`) for UCSF Division of Biostatistics and Bioinformatics (May 3, 2023)

Invited talk at UCSF Biostatistics and Bioinformatics Seminar (May 3, 2023)

Invited talk at UCLA Statistics Seminar (April 27, 2023)

Invited talk at University of Washington Biostatistics Seminar (April 20, 2023)

Invited talk at ITAM Department of Statistics in Mexico (virtual) (March 10, 2023)

Invited guest lecture at Stanford University, Statistics 315A (virtual) (March 8, 2023)

Invited talk at IMS International Conference on Statistics and Data Science, ICSDS in Florence, Italy (December 13, 2022)

Talk at the Iowa State University Graphics Group (October 27, 2022)

Invited talk at USC Division of Biostatistics Seminar (September 29, 2022)

Invited talk at Harvard University Statistics Seminar (September 12, 2022)

Invited talk at Joint Statistical Meetings in Washington DC (August 11, 2022)

Invited talk at University of Washington Statistics Seminar (May 6, 2022)

Invited talk at Robinson College of Business at Georgia State University, Institute for Insight Seminar (virtual) (April 22, 2022)

Invited talk at University of Michigan Statistics Seminar (virtual) (February 25, 2022)

Invited talk at the StatScale (Statistical Scalability for Streaming Data) Seminar - joint Lancaster/Cambridge, UK (virtual) (January 21, 2022)

Invited talk at UCLA Biostatistics Seminar (virtual) (October 6, 2021)

Invited talk at University of Virginia Statistics Colloquium (virtual) (April 23, 2021)

Invited talk at University of North Carolina Chapel Hill STOR (virtual) Colloquium (November 2, 2020)

Talk at the CDC-MIDAS Forecasting Call, "Tools for Comprehensive Evaluation of Forecasters" (September 29, 2020)

Invited talk at (virtual) Joint Statistical Meetings (August 6, 2020)

Invited talk at 2020 Symposium for Data Science and Statistics; e-meeting (June 4, 2020)

Invited talk at CMStatistics conference in London, UK (December 14, 2019)

Invited talk at Tel Aviv University Statistics Seminar (December 10, 2019)

Invited talk at Hebrew University Statistics Seminar (December 9, 2019)

Invited talk at Wharton Statistics Seminar (November 13, 2019)

Invited talk at Two Sigma Investments Academic Seminar Series (October 23, 2019)

Invited talk at University of Toronto Statistics Seminar (September 19, 2019)

Invited talk at Joint Statistical Meetings in Denver, CO (July 31, 2019)

Marketing-DSO Joint Seminar at USC Marshall (March 29, 2019)

Invited talk at Flatiron Institute (Simons Foundation) workshop on "Operator Splitting Methods in Data Analysis" (March 22, 2019)

Invited talk at UC Riverside Statistics Seminar (February 26, 2019)

Invited talk at Maastricht University, D3M (Data-Driven Decision Making) Workshop (December 18, 2018)

Invited talk at CMStatistics conference in Pisa, Italy (December 14, 2018)

Invited talk at UC Santa Barbara Statistics Seminar (December 5, 2018)

Invited talk at INFORMS, Phoenix, AZ (November 5, 2018)

Invited talk (via videoconference) for RIT Data Science Seminar (October 26, 2018)

Invited talk at Cornell Day of Statistics (September 7, 2018)

Invited talk (in topic contributed session) Joint Statistical Meetings, Vancouver, BC, Canada (August 2, 2018)

Invited talk at Statistical Learning and Data Science Conference, New York, NY (June 5, 2018)

Invited talk at Isaac Newton Institute Program on Statistical Scalability, University of Cambridge (April 26, 2018)

Invited talk at SAMSI workshop on “Operator Splitting Methods in Data Analysis” (March 22, 2018)

Invited talk at University of Wisconsin Madison Statistics Seminar (March 14, 2018)

Invited talk at University of Arizona TRIPODS Seminar (March 12, 2018)

Invited talk at University of Arizona Statistics Seminar (March 12, 2018)

Talk at University of Southern California Data Sciences and Operations, PhD Admissions Visit Day (March 1, 2018)

Invited talk at University of Washington Biostatistics Seminar (February 8, 2018)

Invited talk at University of Southern California Probability and Statistics Seminar (February 2, 2018)

Talk at University of Southern California Data Sciences and Operations Brownbag (January 31, 2018)

Invited talk at Texas A&M Statistics Seminar (November 17, 2017)

Poster at “Symposium on Interpretable Machine Learning” at NIPS 2017 in Long Beach, CA (December 7, 2017)

Invited talk at Joint Statistical Meetings in Baltimore, MD (August 2, 2017)

Invited talk at WNAR in Santa Fe, NM (June 27, 2017)

Invited talk at ICSA in Chicago, IL (June 26, 2017)

Invited talk at mini-symposium on “Proximal Techniques for High-Dimensional Statistics” at SIAM Conference on Optimization (May 25, 2017)

Invited talk at Duke Machine Learning Seminar (April 28, 2017)

Invited talk at New England Statistics Symposium (April 22, 2017)

Invited talk at ENAR in Washington, DC (March 14, 2017)

Invited talk at Simons Foundation, New York, NY (January 19, 2017)

Invited talk at University of Chicago Booth School of Business, Econometrics and Statistics Group Seminar (January 12, 2017)

Invited talk at University of Southern California, Marshall School of Business, Statistics Seminar (November 22, 2016)

Invited talk at Binghamton University, Statistics Seminar, Department of Mathematical Sciences (October 13, 2016)

Invited talk at Weill Cornell Medical Campus, Division of General Internal Medicine (October 10, 2016)

Invited talk at Syracuse University, Statistics Seminar, Department of Mathematics (September 27, 2016)

Invited talk (in topic contributed session) Joint Statistical Meetings, Chicago, IL (August 1, 2016)

Invited talk at International Biometric Conference, Victoria, BC, Canada (July 14, 2016)

Talk at useR!2016, Stanford, CA (June 29, 2016)

Poster at ICML 2016 workshop on “Advances in non-convex analysis and optimization,” New York, NY (June 23, 2016)

Invited talk at International Chinese Statistical Association's Applied Statistics Symposium, Atlanta, GA (June 13, 2016)

Invited talk at Statistical Learning and Data Science conference, Chapel Hill, NC (June 7, 2016)

Invited talk at University of Minnesota, Department of Statistics Seminar (April 28, 2016)

Invited talk at ENAR in Austin, TX (March 8, 2016)

"An Introduction to the Lasso" workshop for *Cornell Statistical Consulting Unit* (February 24, 2016)

Invited talk at CFE/CMStatistics conference, London (December 14, 2015)

Invited talk at Cambridge University Statistical Laboratory Seminar (December 11, 2015)

Invited talk at George Mason University, Department of Statistics Seminar (November 13, 2015)

Invited talk at Carnegie Mellon University, Statistical Machine Learning Reading Group (October 21, 2015)

Invited talk at Johns Hopkins University, Department of Biostatistics Seminar (October 19, 2015)

Invited talk at Penn State University, Department of Statistics Seminar (September 3, 2015)

Talk at invited session, Joint Statistical Meetings (August 2015)

Invited talk, Joint 24th ICSA Applied Statistics Symposium and 13th Graybill Conference, session on "Advanced development in big data analytics tools" (June 2015)

Invited talk at Oregon State University, Department of Statistics Seminar (June 2015)

Invited talk at University of Chicago Booth School of Business, Econometrics and Statistics Group Seminar (May 2015)

Invited talk at workshop on "Statistical Inference for Large Scale Data" at Simon Fraser University (April 2015)

Invited talk at Temple University, Department of Statistics Seminar (April 2015)

"An Introduction to the Lasso" workshop for *Cornell Statistical Consulting Unit* (March 15, 2015)

Invited talk at University of Florida, Department of Statistics Seminar (November 2014)

Invited talk as part of a topic contributed session on "Recent Advances in High-Dimensional Statistical Inference" at Joint Statistical Meetings (August 2014)

Invited talk at Data Mining in Business and Industry, session on "Mining for interactions in complex data" (June 2014)

Invited speaker at University of Rochester Biostatistics & Computational Biology Colloquium Series (March 2014)

Talk at Department of Statistical Science Seminar, Cornell University (March 2014).

Invited talk at University of Pittsburgh, Department of Statistics Seminar (March 2014)

Invited talk at SAMSI workshop, "Statistical inference in sparse high-dimensional models: theoretical and computational challenges." (February 2014).

Talk at AI Seminar, Cornell University (February 2014).

Invited talk at Modern Large Scale Statistical Learning Workshop, LA, CA (in honor of Trevor Hastie's 60th Birthday, November 2013).

Invited talk at NYU Stern School of Business Department of Information, Operations & Management Sciences, Statistics Research Seminar (November 2013).

Invited lecture at "International Workshop on the Perspectives on High Dimensional Data Analysis, III" at UBC (May 2013).

Invited talk at Statistical Society of Canada, session on "Statistical Learning" (May 2013).

Two "invited" presentations for 7th graders at Boynton Middle School as part of Math Day (May 2013).

Talk at Cornell Department of Statistical Science Seminar (March 2013).

Invited talk at ISCA 2012, Applied Statistics Symposium, session on "new directions in genomics" (June 2012)

Invited talk at Cornell University, Department of Biological Statistics and Computational Biology Seminar (February 2012)

Invited talk at Carnegie Mellon University, Department of Statistics Seminar (February 2012)

Invited talk at University of Texas at Austin, Division of Statistics + Scientific Computation Seminar (February 2012)

Invited talk at Princeton University, Department of Operations Research and Financial Engineering Seminar (February 2012)

Invited talk at Harvard University, Department of Statistics Seminar (February 2012)

Invited talk at University of Chicago, Department of Statistics Colloquium (January 2012)

Invited talk at University of Southern California, Marshall School of Business, Statistics Seminar (January 2012)

Presented at Stanford Statistics Ph.D. Student Conference, Asilomar, CA (Spring 2011)

Poster presentation at Neural Information Processing Systems (NIPS) 2010 conference, Vancouver, B.C. (Winter 2010)

Contributed talk at Joint Statistical Meetings, Vancouver, B.C. (Summer 2010)

Presented at Western North American Region of the International Biometric Society (WNAR), "Best oral presentation – runner up" in the student paper competition, Seattle, WA (Summer 2010)

Presented at Industrial Affiliates' Day, Stanford Statistics Department (Spring 2010)

Presented at the Molecular Profiling Colloquium, Stanford University (Winter 2010)

Presented at IBM T.J. Watson Research Center "Statistical Machine Learning and its Application" workshop (Fall 2009)

Presented at Stanford Statistics Ph.D. Student Conference, Asilomar, CA (Spring 2009)

## **Selected Professional Activities & Service**

### *PhD & Postdoctoral Advising.*

PhD students advised:

- Farhad De Sousa, (2023–present, USC).

- Sang-wook Lee, joint with Xin Tong (2021–present, USC).
- Greg Faleto (graduated 2023, USC. Next career position: VideoAmp)
- Simeng Shao, joint with Adel Javanmard (graduated 2022, USC. Next career position: Applied Scientist, Amazon)
- Xiaohan Yan (graduated 2018, statistics, Cornell University. Next career position: Data Scientist, Microsoft Corporation)
- Guo Yu (graduated 2018, statistics, Cornell University. Next career position: Post-doctoral Fellow, Department of Statistics, University of Washington. Current position: Assistant Professor of Statistics at UC Santa Barbara)

Postdoctoral scholars advised:

- Louis Christie (2025–present).
- Oh-Ran Kwon (2023–2025). Next career position: Assistant Professor of Statistics at Ohio State University.
- Arkajyoti Saha (2021–2023), at University of Washington (joint with Daniela Witten). Next career position: Assistant Professor of Statistics at UC Irvine.
- Timothy Coleman (2021–2022). Next career position: Google.
- Sangwon (Justin) Hyun (2019–2022). Next career position: Assistant Professor of Statistics at UC Santa Cruz.

*Service to Profession.*

Editorial boards

- Associate Editor of *Journal of the American Statistical Association* (January 2023–present)
- Associate Editor of *Journal of the Royal Statistical Society, Series B* (January 2023–December 2025)
- Associate Editor of *Biometrika* (August 2016–December 2022)
- Associate Editor of *Journal of Computational and Graphical Statistics* (October 2018–December 2022)
- Associate Editor of *Biostatistics* (August 2016–October 2018)

Conference service

- IMS Program Chair for the 2025 WNAR/IMS/JR Annual Meeting (August 2024–June 2025)
- Poster Chair of JSM 2023 (member of program committee in charge of all posters) (August 2022–August 2023)
- Co-chair of the ASA Statistical Learning and Data Science Section Student Paper Competition Committee for JSM 2022 (October 2021–January 2022)
- Organized invited session on “Selective Inference” for the 2022 Annual Meeting of the IMS in London (June 27-30, 2022)
- Program committee for *International Chinese Statistical Association (ICSA) China Conference 2022*
- Program committee for *International Chinese Statistical Association (ICSA) China Conference 2021* (however, conference canceled due to pandemic)
- Chair for Joint Statistical Meetings 2020 invited session on “Ocean Statistical Methodology and Application” (August 6, 2020)

- Scientific committee for *Statistical Learning and Data Science* conference 2020
- Chair for Joint Statistical Meetings 2019 invited session on “Recent Advances in Multiple Testing and False Discovery Rate Analysis” (July 30, 2019)
- Judge for Joint Statistical Meetings 2019 Statistical Learning and Data Science student poster session (July 29, 2019)
- Organized invited session on “Computational Statistics for Large-Scale Biological Data” for Symposium on Data Science and Statistics 2019 [Beyond Big Data: Building Data Tools] (May 29–June 1, 2019)
- Chair for Joint Statistical Meetings 2018 Introductory Overview Lecture on “Reproducibility, Efficient Workflows, and Rich Environments,” Vancouver, BC, Canada (July 31, 2018)
- Chair for Joint Statistical Meetings 2017 SPEED poster session on “Reliable Statistical Learning and Data Science.” (August 1, 2017)
- Judge for New England Statistics Symposium poster competition (April 2017)
- Scientific committee for *Statistical Learning and Data Science* conference 2018
- Chaired session at CM Statistics in London (December 2015)
- Organized invited session on “Statistical Learning Using Convex Optimization” at Joint Statistical Meetings 2015 (August 2015)
- Chaired session on “Time Series Data Mining” at Joint Statistical Meetings 2014 (August 2014)

Member of Advisory Board for Centre for Doctoral Training in Statistics and Operational Research in partnership with Industry (STOR-i), Lancaster University (August 2024–present)

Stakeholder for CMU Masters of Statistical Practice Capstone Project (Spring 2022)

NSF review panelist (2017)

Referee for *Annals of Statistics*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society: Series B*, *Biometrika*, *NIPS* conference, *Annals of Applied Statistics*, *Journal of Computational and Graphical Statistics*, *Journal of Machine Learning Research*, *Technometrics*, *Electronic Journal of Statistics*, *Journal of Multivariate Analysis*, *Biometrics*, *Biostatistics*, *Journal of Statistical Planning and Inference*, *Computational Statistics and Data Analysis*, *BMC Bioinformatics*, *Statistics and Computing*, *Statistical Modelling*, *Machine Learning*, *Canadian Journal of Statistics*.

External member of PhD committees at other universities

- External committee member for Ludwig Maximilian University PhD student in statistics (defense, July 30, 2024)
- External committee member for UCLouvain PhD student in statistics (private defense, September 5, 2023; public defense, November 7, 2023)
- External committee member for NCSU PhD student in statistics (May 13, 2022; October 3, 2022)
- External committee member for NYU PhD student in microbiology and computational biology (December 12, 2016)

### *University Service.*

*At USC (2017–present):*

Organizer of industry panel at DSO’s SEEDS conference (January 2024)

DSO research committee (starting Fall 2023)

DSO representative of Marshall *Research and Faculty Recognition Committee* (August 2022–present)

Co-mentor for two students as part of DSO's *Summer Scholars* program (Summer 2025)

Advised USC undergraduate student in project (Summer 2022–Fall 2022) as part of Fall 2022 Provost's Research Fellowship

Proposed and taught two-day, welcome-week micro-seminar on statistics for incoming USC students across university (August 18–19, 2022)

Interim graduate program coordinator for Data Sciences and Operations and also more specifically for the Statistics PhD Program (Spring 2020)

See “PhD & Postdoctoral Advising” section above (not repeated here)

Committee member (non-advisor) for PhD students at USC

- External committee member for PhD dissertation committee (biostatistics; USC; qualifier May, 2025)
- Committee member for PhD qualifier (statistics; Marshall; Nov 6, 2023)
- External committee member for PhD dissertation committee (mathematics; USC; qualifier November 13, 2023; defense July 21, 2025)
- External committee member for PhD dissertation committee (computer science; USC; proposal January 12, 2023; defense May 8, 2023)
- Committee member for PhD (statistics; Marshall; qualifier February 6, 2023; defense June 2, 2025)
- External committee member for (biostatistics; USC; qualifier April 13, 2020; defense July 1, 2021)
- External committee member for PhD qualifier (oceanography; August 4, 2020; pre-meeting: February 5, 2020; defense: July 10, 2024)
- Committee member for PhD qualifier (statistics; Marshall; November 20, 2019)
- External committee member for PhD qualifier (oceanography; USC; October 19, 2018)
- Committee member for PhD defense (statistics; Marshall; July 27, 2018)

Faculty speaker at DSO 621 for Marshall PhD students (September 8, 2022; October 4, 2019; October 6, 2017)

Proposed and taught two-day, welcome-week micro-seminar on statistics for incoming USC students across university (August 22–23, 2019)

DSO Statistics PhD curriculum committee member (September 2018)

Faculty speaker for DSO (Statistics) Prospective PhD Visit Day (March 1, 2018)

Faculty attendee/participant at Marshall PhD Recruiting Fair, answering questions from prospective applicants interested in DSO (November 7, 2017)

*At Cornell (2012–2017):*

Faculty consultant for *Cornell Statistical Consulting Unit* (Fall 2012–Spring 2017)

Academic advisor to Biometry & Statistics majors; e.g., advising 20+ students in 2016. (Fall 2013–Spring 2017)



Advisor to Masters of Professional Statistics students, e.g. advising 4 students in 2016. (Fall 2013–Spring 2017)

Graduate committee member for 12 Cornell PhD students (11 completed).

Department of Statistical Science seminar coordinator (Fall 2015–Spring 2017)

BSCB representative to faculty senate of College of Agriculture and Life Sciences (Fall 2014–Spring 2015)

Google *Summer of Code* co-mentor (with David Matteson) to Will Nicholson (Summer 2014)

Principal Investigator for Department of Statistical Science's parallel computing cluster (Fall 2012–Spring 2017).

*At Stanford (2007–2012):*

Coordinator for Hastie-Taylor-Tibshirani weekly group meeting (Fall 2010–Spring 2011)

Stanford Statistics Department's Student Reader for the Applied Statistics Qualifying Exam (Summer 2009)

Chair of the Statistical Learning Session of the Stanford Statistics Ph.D. Student Conference, Asilomar, CA (Spring 2009)

*Service to Community.*

Member of board of directors of *Wonderfest: The Bay Area Beacon of Science*, a non-profit focused on the popularization of science (Fall 2011–present)

Mentor for two students as part of USC's *JumpStart Program*, which "aims to provide a pathway to PhD programs" for community college and undergraduate students (Summer 2024)

Public outreach talk through Wonderfest at Berkeley Public Library (July 13, 2024)

CMU Delphi COVID-19 Response Team, Core Member (April 27, 2020–2022)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," e-workshop (April 19, 2025)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," e-workshop (December 30, 2023)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," e-workshop (March 18, 2023)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," e-workshop (March 19, 2022)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," e-workshop (November 21, 2020)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," e-workshop (May 16, 2020)

Public outreach talk on statistics through Wonderfest at Blue River Technology, Sunnyvale, CA (November 19, 2019)

Public outreach talk on statistics through Wonderfest at the Alameda Free Library, Alameda, CA (June 16, 2019)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," in San Francisco (May 18, 2019)

Public outreach talk through Wonderfest at HopMonk Tavern, Novato, CA (March 25, 2019)

Workshop for Wonderfest's Science Envoy Program, "Beyond PowerPoint: Powerful Points about Visual Aids," at UC Berkeley (March 10, 2018)

Ithaca Sciencenter Showtime! presentation - public outreach on statistics. "Card Shuffling, Coin Flipping, and Coincidences. Thinking Like a Statistician." (March 25, 2017)

Ithaca Sciencenter Showtime! presentation, "The Statistics Shuffle" - public outreach on statistics. "Explore the power of statistics through card shuffling with Jacob Bien, Assistant Professor of Biological Statistics and Computational Biology at Cornell University." (August 27, 2016)

January 12, 2026