

October 2024

Li Ma

Department of Statistical Science
Duke University Box 90251
Durham, NC 27708-0251

Phone: (919) 684-2871
Email: li.ma@duke.edu
Web: <https://lm186.github.io>

Education

Stanford University, Stanford, CA, September 2006 – June 2011

Ph.D. in Statistics. Advisor: Wing Hung Wong

University of Chicago, Chicago, IL, September 2002 - August 2006

M.S./B.A. in Statistics. Advisor: Michael L. Stein

B.A in Mathematics, and **B.A.** in Economics.

Research Interests

- Methodology: generative models, tree-based models and methods, nonparametric methods, multi-scale inference, Bayesian modeling, scalable methods for massive data.
- Applications: microbiome analysis, computational immunology, genetic epidemiology, functional genomics, and other applied problems involving big data.

Professional Experience

Professor 07/2023 -
Department of Statistical Science, Duke University, Durham, NC

Professor (secondary) 12/2023 -
Department of Biostatistics and Bioinformatics, Duke University Medical School, Durham, NC

Associate Professor 07/2018 - 06/2023
Department of Statistical Science, Duke University, Durham, NC

Associate Professor (secondary) 07/2020 - 11/2023
Department of Biostatistics and Bioinformatics, Duke University Medical School, Durham, NC

Assistant Professor 07/2011 - 06/2018
Department of Statistical Science, Duke University, Durham, NC

Visiting Professor 01/2022 - 06/2022
Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy

Research Fellow and Working Group Leader 08/2013 - 05/2014, 08/2018 - 05/2019
Statistical and Applied Mathematical Sciences Institute, Durham, NC

Visiting Scholar 03/2016 - 06/2016
Department of Statistics, University of Chicago, Chicago, IL

Decision support analyst (statistician) intern

Google, Inc., Mountain View, CA

06/2009 - 09/2009

Funding

Ongoing

1. **NIH 1R01GM135440.** Role: PI (sole).

Start/End: September 2020 - August 2025. Total cost: \$1,717,909.

Title: Statistical modeling of cross-sample variation and learning of latent structures in microbiome sequencing data.

2. **NSF DMS-2152999.** Role: PI (MPI with Pulong Ma at Iowa State University).

Start/End: August 2022 - July 2025. Total cost (Duke): \$120,000.

Title: Collaborative research: Bayesian residual learning and random recursive partitioning methods for Gaussian process modeling.

3. **NSF EEC-2133504.** Role: Research Thrust Lead (PI: Gunsch)

Start/End: September 2022 - August 2027. Total cost: \$26,000,000.

Title: NSF Engineering Research Center for Precision Microbiome Engineering (PreMiEr).

4. **NIH 1R01-DK138121.** Role: Co-Investigator (PI: Siddiqui)

Start/End: February 2024 - November 2028. Total cost: \$2,655,690.

Title: Predicting phenotypes in benign urology.

Completed

5. **NSF DMS-2013930.** Role: PI (sole).

Start/End: July 2020 - June 2024. Total cost: \$200,000.

Title: Advances in Bayesian nonparametric methods for jointly modeling multiple data sets.

6. **NSF CAREER Award DMS-1748789.** Role: PI (sole).

Start/End: September 2018 - August 2024. Total cost: \$400,000.

Title: CAREER: Advances in multi-scale Bayesian inference and learning on massive data.

7. **NSF DMS-1309057.** Role: PI (sole).

Start/End: July 2013 - June 2016. Total cost: \$159,873.

Title: Bayesian recursive partitioning and inference on the structure of high-dimensional distributions.

8. **NSF DMS-1612889.** Role: PI (sole).

Start/End: July 2016 - June 2020. Total cost: \$345,116.

Title: Graphical multi-resolution scanning for cross-sample variations.

9. **Google Faculty Research Award.** Role: PI (sole).

Start/End: July 2016 - June 2017. Direct cost: \$44,840.

Title: Graphical multi-resolution scanning for cross-sample variations.

10. **NIH 5R03-AG060082.** Role: Co-Investigator. (PI: Siddiqui).

Start/End: July 2018 - June 2020.

Title: Effects of aging and the urinary microbiome on recurrent urinary tract infections.

11. **Duke Microbiome Center Development Grant.** Role: Co-Investigator (PI: Sung).
Start/End: July 2020 - June 2021.
Project title: A randomized, double-blind, placebo-controlled trial of the probiotics to eliminate COVID-19 transmission in exposed household contacts (PROTECT-EHC).
12. **NSF DMS-1938935.** Role: PI (conference).
Start/End: January 2020 - December 2021. Direct cost: \$30,000.
Title: ISBA 2020: 15th World Meeting of the International Society for Bayesian Analysis – June 29-July 3, 2020.
13. **Translating Duke Health Pilot Grant.** Role: PI (joint: Chan & Ma).
Start/End: July 2020 - February 2022. Total cost: \$100,000.
Title: Statistical methods for comparative analysis of multi-sample flow cytometry data.
14. **AAOGF Bridge Grant.** Role: Co-Investigator (PI: Siddiqui).
Start/End: October 2020 - June 2022.
Title: Urinary Lactobacilli and microbial interference in the aging urinary microbiome.

Awards & Honors

Fellow of the International Society for Bayesian Analysis	2024
Fellow of the American Statistical Association	2024
ISBA Certificate of Appreciation for Outstanding Service	2021
NSF CAREER Award	2018
Google Faculty Research Award in Machine Learning	2016
Larry Yung Stanford Interdisciplinary Graduate Fellowship	2009 - 2011
Gerhard Casper Stanford Graduate Fellowship	2006 - 2009
Student Marshal, University of Chicago	2005
Phi Beta Kappa	2005

Publications (* indicates student and ** indicates postdoc advisees)

Peer-reviewed articles

1. Ma, L., Stein, M.L., Wang, M., Shelton, A.O., Pfister, C.A., and Wilder, K.J. (2010). A method for unbiased estimation of population abundance along curvy margins. *Environmetrics*. Vol. 22, No. 3, 330-339
2. Ma, L., Assimes, T., Asadi, N.B., Iribarren, C., Quertermous, T., and Wong, W.H. (2010). An “almost exhaustive” search based permutation method for detecting epistasis in disease association studies. *Genetic Epidemiology*. Vol. 34, No.5, 434–443.
3. Ma, L., Mease, D., and Russell, D. (2010). A four group cross-over design for measuring irreversible treatments on web search tasks. *Proceedings of the 44th Hawaiian International Conference on System Sciences*.
4. Wong, W.H. and Ma, L. (2010). Optional Pólya tree and Bayesian inference. *Annals of Statistics*. Vol. 38, No. 3, 1433–1459.
5. Ma, L. and Wong, W.H. (2011). Coupling optional Pólya trees and the two sample problem. *Journal of the American Statistical Association. T&M*. Vol. 106, No. 496, 1553-1565.

6. Ma, L., Wong W.H. and Owen A.B. (2012). A sparse transmission disequilibrium test for haplotypes based on Bradley-Terry graphs. *Human Heredity*. Vol. 73, No. 1, 52-61.
7. Ma, L. (2013). Adaptive testing of conditional association through recursive mixture modeling. *Journal of the American Statistical Association*. *T&M*. Vol. 108, No. 504, 1493-1505.
8. Ma L. (2015). Scalable Bayesian model averaging through local information propagation. *Journal of the American Statistical Association*. *T&M*. Vol. 110, No. 510, 795-809.
9. Soriano J* and Ma L. (2017). Probabilistic multi-resolution scanning for two-sample differences. *Journal of the Royal Statistical Society. Series B*. Vol. 79, No. 2, 547-572.
10. Ma L. (2017). Adaptive shrinkage in Pólya tree type models. *Bayesian Analysis*. Vol. 12, No. 3, 779-805. (Featured in editor's invited session "Highlights from Bayesian Analysis" at JSM 2017.)
11. Ma L. (2017). Recursive partitioning and multi-scale modeling on conditional densities. *Electronic Journal of Statistics*. Vol. 11, No. 1, 1297-1325.
12. Ma L and Soriano J*. (2018). Efficient functional ANOVA through wavelet-domain Markov groves. *Journal of the American Statistical Association*. *T&M*. Vol. 113, No. 3, 802-818.
13. Tang Y, Ma L, and Nicolae D.L. (2018). A phylogenetic scan test on Dirichlet-tree multinomial model for microbiome data. *Annals of Applied Statistics*. Vol. 12, No. 1, 1-26.
14. Ma L and Soriano J*. (2018). Analysis of distributional variation through graphical multi-scale Beta-Binomial modeling. *Journal of Computational and Graphical Statistics* Vol. 27, No. 3, 529-541.
15. Ma L and Mao J*. (2019). Fisher exact scanning for dependency. *Journal of the American Statistical Association*, *T&M*. Vol. 114, No. 525, 245-258.
16. Soriano J* and Ma L. (2019). Mixture modeling on related samples through ψ -stick breaking and kernel perturbation. *Bayesian Analysis*. Vol. 14, No.1, 161-180.
17. Mao J*, Chen Y*, and Ma L. (2020). Bayesian graphical compositional regression for microbiome data. *Journal of the American Statistical Association*, *A&CS*. Vol. 115, No. 530, 610-624.
18. Christensen J* and Ma L. (2020). A Bayesian hierarchical model for related densities using Pólya trees. *Journal of the Royal Statistical Society, Series B*. Vol. 82, 127-153.
19. Liu R, Li M, and Ma L. (2020). CARP: Compression through adaptive recursive partitioning for multi-dimensional images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14306-14314.
20. Giri V, Kegerreis K, Ren Y, Bohannon L, Lobaugh-Jin E, Messina J, Matthews A, Mowery Y, Sito E, Lassiter M, Saullo J, Jung S, Ma L, Greenberg M*, Andermann T, van den Brink M, Peled J, Gomes A, Choi T, Gasparetto C, Horwitz M, Long G, Lopez R, Rizzieri D, Sarantopoulos S, Chao N, Allen D, and Sung A. (2021). Chlorhexidine gluconate bathing reduces the incidence of bloodstream infections in adults undergoing inpatient hematopoietic cell transplantation. *Transplantation and Cellular Therapy*. Vol. 27, No. 1, 262e1-262e11.

21. Ramalingam S, Siamakpour-Reihani S, Bohannon L, Ren Y, Sibley, Sheng J, Ma L, Nixon AB, Lyu J, Parker DC, Bain B, Muehlbauer M, Ilkayeva O, Kraus VB, Huebner J, Spitzer T, Brown J, Peled J, van den Brink M, Gomes A, Choi T, Gasparetto C, Horwitz M, Long G, Lopez R, Rizzieri D, Sarantopoulos S, Chao N, and Sung AD. (2021). A phase 2 trial of the somatostatin analog pasireotide to prevent GI toxicity and acute GVHD in allogeneic hematopoietic stem cell transplant. *PLOS ONE*. Vol. 16, No. 6.
22. Vaughan M, Mao J*, Karstens L, Ma L, Amundsen C, Schmader K, Siddiqui N. (2021). The urinary microbiome in postmenopausal women with recurrent urinary tract infections. *Journal of Urology*. Vol. 206, No. 5, 1222-1231.
23. Li M and Ma L. (2022). Learning asymmetric and local features in multi-dimensional data through wavelets with adaptive recursive partitioning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 44, No. 11, 7674-7687.
24. Mao J* and Ma L. (2022) Dirichlet-tree multinomial mixtures for clustering microbiome compositions. *Annals of Applied Statistics*. Vol. 16, No. 3, 1476-1499.
25. Gorsky S* and Ma L. (2022) Multiscale Fisher's independence test for multivariate dependence (with discussion). *Biometrika*. Vol. 109, No. 3, 569-587.
26. Luo K, Zhong J, Safi A, Hong L.K., Tewari A.K., Song L, Reddy T.E., Ma L, Crawford G.E., Hartemink A.J. (2022). Quantitative occupancy of myriad transcription factors from one DNase experiment enables efficient comparisons across conditions. *Genome Research*. 32: 1183-1198.
27. Siddiqui N, Ma L, Brubaker L, Mao J*, Hoffman C, Wang Z*, Karstens L. (2022) Updating urinary microbiome analyses to enhance biologic interpretation. *Frontiers in Cellular and Infection Microbiology*. 12:789439.
28. Vaughan M, Zemtsov G.E., Dahl E.M., Karstens L, Ma L, Siddiqui N. (2022) Concordance of urinary microbiota detected by 16S rRNA amplicon sequencing versus expanded quantitative urine culture. *American Journal of Obstetrics & Gynecology*. Vol. 227, No. 5, 773-775.
29. LeBlanc P* and Ma L. (2023) Microbiome subcommunity learning with logistic-tree normal latent Dirichlet allocation. *Biometrics*. Vol. 79, Iss.3, 2321-2332.
30. Ji Z and Ma L. (2023) Controlling taxa abundance improves metatranscriptomics differential analysis. *BMC Microbiology*. 23:60.
31. Awaya N* and Ma L. (2024) Hidden Markov Pólya trees for high-dimensional distributions. *Journal of the American Statistical Association, T&M*. Vol.119, No.545, 189-201.
32. Gorsky S*, Chan C, and Ma L. (2024) Coarsened mixtures of hierarchical skew normal kernel for flow cytometry analyses. *Bayesian Analysis*. Vol.19, No.2, 439-463.
33. Liu R, Li M, and Ma L. (2024). Efficient in-situ image and video compression through probabilistic image representation. *Signal Processing*. Vol. 215, 109268.
34. Liu L and Ma L. (2024) Spatial adaptation by Bayesian unsupervised trees. *Proceedings of Thirty Seventh Conference on Learning Theory, PMLR*. 247:3556-3581.

35. Awaya N* and Ma L. (2024) Unsupervised tree boosting for learning probability distributions. *Journal of Machine Learning Research*. 25(198):1-52.
36. Horiguchi A**, Chan C, and Ma L. (2024) A tree perspective on stick-breaking models in covariate-dependent mixtures. *To appear in Bayesian Analysis*.

Other publications

37. Ma L. (2019). Discussion on “Latent Nested Nonparametric Priors” by Camerlanghi et al. *Bayesian Analysis*. Vol. 14, No. 4, 1303-1356.
38. Gorsky S* and Ma L. (2022). Rejoinder: “Multiscale Fisher’s independence test for multivariate dependence”. *Biometrika*. Vol. 109, No. 3, 605–609.

Submitted manuscripts

39. Wang Z*, Mao J*, and Ma L. (2024) Differential abundance analysis for microbiome compositions with logistic-tree normal models. *Submitted*.
40. Horiguchi A**, Ma L, and Szabo B. (2024) Sampling depth trade-off in function estimation under a two-level design. *Submitted*.
41. Siddiqui NY, Diaz DM, Wang Z*, Rouhier A, Ma L, and Karstens L. (2024) Impact of menopause and vaginal estrogen on the human female urinary microbiome. *Submitted*.
42. Luo H, Horiguchi A**, and Ma L. (2024) Efficient decision trees for tensor regressions. *Submitted*.
43. Wang Z*, Awaya N*, and Ma L. (2024) Generative modeling of conditional densities through additive tree flows. *Submitted*.
44. Wei G and Ma L. (2024) Stream-level flow matching from a Bayesian decision theoretic perspective. *Submitted*.

Published software

R packages: 4 on CRAN—MRS, grove, MultiFIT, and COMIX; 12 on Github—PTT, LIPS, remcat, FES, BGCR, HAPT-package, WARP, DTMM, CARP, SMCMP, LTNLDA and boostPM.

Patents

1. Wong W.H. and Ma L. (2012) *US patent 2012-0078821*.
Title: Methods for unsupervised learning using optional Pólya trees and Bayesian inference.

Conference Presentations

1. Contributed poster – 18th Annual Meeting of the International Genetic Epidemiology Society. Kahuku, HI. (2009)
2. Invited talk – 42nd Symposium on the Interface of Computing Science and Statistics. Cary, NC. (2011)
3. Contributed poster – ISBA 2012, Kyoto, Japan. (2012)

4. Invited talk – Joint Statistical Meeting 2012. San Diego, CA. (2012)
5. Invited talk – 5th International Conference of the ERCIM on Computing & Statistics. Oviedo, Spain. (2012)
6. Invited talk – International Workshop on Bayesian Model Selection. Shanghai, China. (2013)
7. Invited talk & contributed poster – 2013 ENAR Spring Meeting. Orlando, FL. (2013)
8. Invited talk – The IMS-China International Conference on Statistics and Probability. Chengdu, China. (2013)
9. Contributed talk (by student) – Joint Statistical Meeting 2013. Montreal, Canada. (2013)
10. Invited talk – 6th International Conference of the ERCIM on Computing & Statistics. London, UK. (2013)
11. Invited discussant – OBayes 250. Durham, NC. (2013)
12. Invited talk – The 6th International Statistics Forum at Renmin University. Beijing, China. (2014)
13. Invited talk – 2014 International Workshop on Controlling Multiplicity in Statistical Analysis. Shanghai, China. (2014)
14. Invited talk – 2014 Joint Statistical Meeting. Boston, MA. (2014)
15. Invited talk – 12th ISBA World Meeting. Cancun, Mexico. (2014)
16. Invited talk – 7th International Conference of the ERCIM on Computing & Statistics. Pisa, Italy. (2014)
17. Invited talk – The 29th New England Statistical Symposium. Storrs, CT. (2015)
18. Invited talk – 10th Conference on Bayesian Nonparametrics. Raleigh, NC. (2015)
19. Contributed talk (by student)– 10th Conference on Bayesian Nonparametrics. Raleigh, NC. (2015)
20. Invited talk – SAMSI workshop on Bayesian Nonparametrics. Raleigh, NC. (2015)
21. Invited talk – 8th International Conference of the ERCIM on Computing & Statistics. London, UK. (2015)
22. Invited talk – 13th ISBA World Meeting. Sardinia, Italy. (2016)
23. Contributed poster (by student) – 13th ISBA World Meeting. Sardinia, Italy. (2016)
24. Invited talk – 2016 Joint Statistical Meeting. Chicago, IL. (2016)
25. Invited talk – International Conference on Advances in Interdisciplinary Statistics and Combinatorics. Greensboro, NC. (2016)

26. Invited talk – 10th International Conference of the International Chinese Statistical Association. Shanghai, China. (2016)
27. Invited talk – The 31st New England Statistical Symposium. Storrs, CT. (2017)
28. Invited talk – 11th Conference on Bayesian Nonparametrics. Paris, France. (2017)
29. Invited talk – 2017 Joint Statistical Meeting. Baltimore, MD. (2017)
30. Invited talk – BIRS-CMO Workshop on Bayesian nonparametrics. Oaxaca, Mexico. (2017)
31. Invited discussant – 2017 Objective Bayes Meeting. Austin, TX. (2017)
32. Selected participant – 2017 NIH-BD2K Innovation Lab on Quantitative Approaches to Biomedical Data Science Challenges in our Understanding of the Microbiome. Beverly, MA. (2017)
33. Invited talk – ENAR Spring Meeting. Atlanta, GA. (2018)
34. Invited talk – 14th ISBA World Meeting. Edinburgh, UK. (2018)
35. Invited talk – 2018 ICSA Applied Statistics Symposium. Rutgers, NJ. (2018)
36. Invited talk – 2018 Joint Statistical Meeting. Vancouver, BC, Canada. (2018)
37. Invited talk – 11th International Conference of the ERCIM WG on Computational and Methodological Statistics. Pisa, Italy. (2018)
38. Invited talk – 2019 ICSA Applied Statistics Symposium. Raleigh, NC. (2019)
39. Invited talk – JSM 2020. Online Virtual Conference. (2020)
40. Contributed poster – CVPR 2020. Online Virtual Conference. (2020)
41. Invited discussant – BAYSM:O. Online Virtual Conference (2020)
42. Invited speaker – 45th Annual Spring Lecture Series at University of Arkansas. Online Virtual Conference. (2020)
43. Invited talk – 5th EAC-ISBA Conference. Online. (2021)
44. Invited discussant – BAYSM 2021. Online. (2021)
45. Invited talk – JSM 2021. Online. (2021)
46. Invited talk – IMSI Workshop on “Multiscale Microbial Communities – Dynamic Models, Ecology and One Health”. Chicago, IL. (2022)
47. Invited talk – ISBA 2022. Montreal, Canada. (2022)
48. Invited talk – JSM 2022. Washington, DC. (2022)
49. Invited talk – 15th International Conference of the ERCIM WG on Computational and Methodological Statistics. London, UK. (2022)

50. Invited talk – 2023 ICSA Applied Statistics Symposium, Ann-Arbor, MI. (2023)
51. Invited talk – Workshop on "Approximation Methods in Bayesian Analysis". Marseille, France. (2023)
52. Invited talk – JSM 2023. Toronto, Canada. (2023)
53. Invited talk – Statistics and Data Science in Biomedicine. Stanford, CA. (2023)
54. Invited participant – NUS-IMS workshop on Interpretable Inference via Principled BNP Approaches in Biomedical Research and Beyond. Singapore. (2024)
55. Invited mini-course – 3rd BNP-ISBA Networking Workshop. Singapore. (2024)
56. Invited talk – 3rd IMS International Conference on Statistics and Data Science. Nice, France. (2024)
57. Invited talk – ENAR Spring Meeting, New Orleans, LA. (2025)
58. Keynote talk – 14th International Conference on Bayesian Nonparametrics, Los Angeles, CA. (2025)

Departmental Seminars

1. Statistics seminar, University of Chicago (2010)
2. Biostatistics workshop, Stanford University (2010)
3. Statistics seminar, University of Florida (2011)
4. Statistics seminar, Rutgers University (2011)
5. Statistics seminar, University of Wisconsin at Madison (2011)
6. Operations research and financial engineering seminar, Princeton University (2011)
7. Statistics seminar, University of California at Berkeley (2011)
8. Statistics seminar, Ohio State University (2011)
9. Statistics seminar, University of Chicago (2011)
10. Statistics seminar, Columbia University (2011)
11. Statistics seminar, Duke University (2011)
12. Biostatistics workshop, Stanford University (2011)
13. Statistics colloquium, University of South Carolina (2013)
14. Statistics seminar, Pennsylvania State University (2013)
15. Statistics seminar, Collegio Carlo Alberto (2013)
16. Statistics seminar, Clemson University (2014)

17. Statistics seminar, University of Chicago (2014)
18. Statistics colloquium, Virginia Tech (2015)
19. Statistics seminar, UCLA (2015)
20. Biostatistics seminar, University of Florida (2015)
21. Bayesian statistics seminar, NC State University (2015)
22. Statistics colloquium, Purdue University (2016)
23. Econometrics and statistics colloquium, University of Chicago Booth School of Business (2016)
24. Statistics seminar, University of Pennsylvania Wharton School of Business (2016)
25. Statistics seminar, Columbia University (2017)
26. Statistics seminar, University of Michigan (2018)
27. Biostatistics and Medical Informatics seminar, University of Wisconsin at Madison (2018)
28. Biostatistics seminar, Yale University (2019)
29. Statistics Summer School, Shandong University (2021)
30. Statistics seminar, Texas A&M University (2021)
31. Statistics seminar, Clemson University (2021)
32. Statistics seminar, Bocconi University (2022)
33. Statistics seminar, University of Milan Bicocca (2022)
34. Statistics seminar, Duke University (2022)
35. Statistics colloquium, University of Chicago (2024).
36. Statistics seminar, University of Pittsburgh (2024).
37. Statistics seminar, University of North Carolina Chapel Hill (2024).
38. Statistics colloquium, University of Central Florida (2024).

Professional Memberships

American Statistical Association
Institute of Mathematical Statistics
International Society for Bayesian Analysis
International Chinese Statistical Association

Teaching at Duke

STA 602: Bayesian Statistical Modeling and Data Analysis Fall 2020, 2021, 2022, 2023

STA 702: Bayesian Statistical Modeling and Data Analysis	Fall 2024
STA 732: Theory of Statistics	Spring 2015, 2017, 2018, 2019, 2020, 2021, 2024
STA 841: Methods for Categorical Data	Fall 2011, 2012, 2013, 2014, 2015, 2016
STA 250: Intro to Mathematical Statistics	Spring 2012, 2014, Fall 2014, 2017
STA 641: Nonparametric models and learning (1/3)	Spring 2020
STA 613: Statistical methods for computational biology (1/2)	Spring 2019
STA 701: Readings in statistics literature	Fall 2015, 2016, Spring 2017, Fall 2024
STA 790: Statistical methods for high-dimensional data	Fall 2012
STA 790: Bayesian nonparametric models and methods (1/3)	Spring 2021, Fall 2023
STA 941: Nonparametric models and learning (1/3)	Fall 2022

Other Teaching

41007 (Bocconi): Statistical modeling and learning with trees and ensembles Spring 2022

Departmental and University Service

1. DSS Promotion Review Committee (Chair), 2024.
2. Interim Director of the Master's in Statistical Science Program, 2024.
3. NSF Engineering Research Center for Precision Microbiome Engineering — Research Thrust Lead, 2022 –
4. Duke University Masters Advisory Council, 2022-2023.
5. DSS PhD admissions—Committee 2011-2012, 2016-2017, 2017-2018, 2018-2019; ad-hoc application review and phone interview 2013, 2014, 2015.
6. DSS MS admissions committee, 2014, 2020, 2021, 2024.
7. DSS Masters Advisory Committee, 2021, 2023. Chair, 2024 -
8. DSS Undergraduate Curriculum Committee, 2019-2021.
9. DSS Tenure-track Search Committee, 2016-2017, 2017-2018.
10. DSS Reappointment Committee, 2018-2019, 2020-2021, 2024.
11. DataFest: Workshop instructor (2017), judge (2012, 2013, 2015, 2018), consultant (2014).
12. DSS Computing Committee, 2013–2019.
13. DSS BEST Award Committee, 2014 - 2021, 2023 -.

14. DSS PhD First Year Exam coordination: 2014 (coordinator), 2015 (coordinator), 2021 (committee member), 2022 (committee chair).
15. DSS Mentorship Committee, 2012.
16. DSS Seminar Chair: 2014, 2015.
17. Presenter for Duke Open House Math, Statistics, and Computer Science panel (2015).
18. Duke Datathon Judge: 2020, 2021.
19. CBB PhD admissions—ad-hoc interviewer, 2018, 2019, 2021.

Professional Service

1. Referee for various professional journals:
American Journal of Human Genetics, Annals of Applied Statistics, Annals of Statistics, Bayesian Analysis, Biometrics, Biometrika, Biostatistics, BMC Bioinformatics, Canadian Journal of Statistics, Communications in Statistics, Computational Statistics and Data Analysis, Electronic Journal of Statistics, Genome Biology, IEEE/ACM Transactions on Computational Biology and IEEE Transactions on Signal Processing, Bioinformatics, Journal of Computational and Graphical Statistics, Journal of the American Statistical Association, Journal of the Royal Statistical Society - Series B, Journal of the Royal Statistical Society - Series C, Journal of Statistical Computation and Simulation, PLoS ONE, Proceedings of the National Academy of Sciences, Scandinavian Journal of Statistics, STAT, Statistical Science, Statistica Sinica, Technometrics, The American Statistician.
2. Program committee: Joint Statistical Meeting 2018, Vancouver, Canada, 2017-2018;
 12th Conference on Bayesian Nonparametrics, Oxford, UK, 2017-2019;
 Chair, 2021 ISBA World Meeting, Online, 2018-2021;
 2023 ICSA Applied Statistics Symposium, Ann Arbor, MI, 2022-2023.
3. International Society for Bayesian Analysis: Program Council, 2018 - 2020 (Chair 2019);
 Secretary for the BNP Section, 2015 - 2016; Savage Award Committee, 2021 - 2022;
 Lindley Prize Committee, 2021 - 2022; Blackwell-Rosenbluth Award Committee, 2022 and 2024;
 Board of Directors (elected), 2022 - 2025; Nomination Committee, 2024.
4. Associate Editor: Journal of the American Statistical Association - T&M, 2021 - ;
 Bayesian Analysis, 2019 - ;
 Journal of Business and Economic Statistics, 2024 - ;
 Electronic Journal of Statistics Special Issue on Bayesian Nonparametrics,
 2015 - 2016.
5. Grant review: Panelist for NSF 2012, 2017, 2019, 2023, 2024. Reviewer for NSA 2016.
 Panelist for NIH 2021, 2022. Reviewer for RGC of Hong Kong in 2021. Reviewer for MIUR of Italy in 2021. Reviewer for ISF of Israel in 2023 and 2024.
6. Group leader: SAMSI working group on statistical genetics and genomics, 2013-2014.
 SAMSI working group on observational microbiome data, 2018-2019.
7. Discussion leader/moderator: SAMSI Workshop on Bayesian nonparametrics, 2015.

8. Organizing committee: 10th Conference on Bayesian Nonparametrics. Raleigh, NC, 2015.
10. Session organizer: ERCIM 2013, 2014, 2015, 2016, 2018. JSM 2014, 2016, 2018, 2019. ISBA 2016. ENAR spring meeting 2018. ICSA Applied Statistics Symposium 2023.
11. Session chair: ENAR spring meeting 2013, IMS-China meeting 2013, ERCIM 2013, ERCIM 2014, ERCIM 2015, ISBA 2018, CMStatistics 2018, JSM 2020, JSM 2021, JSM 2023.
12. External reviewer for tenure and promotion for various universities.

Mentoring

Ph.D. dissertation advisor:

1. Jacopo Soriano (DSS, 06/2012 – 05/2015 degree completed)
2. Jonathan Christensen (DSS, 06/2013 – 12/2017 degree completed)
3. Shai Gorsky (DSS, 05/2017 – 05/2020 degree completed)
4. Jialiang Mao (DSS, 07/2016 – 08/2020 degree completed)
5. Naoki Awaya (DSS, 05/2019 – 08/2022 degree completed)
6. Patric LeBlanc (DSS, 05/2019 – 05/2023 degree completed, co-advised with David Banks)
7. Zhuoqun Wang (DSS, 05/2020 –)
8. Benedetta Bruni (DSS, 01/2024 – , co-advised with Simon Mak)

Postdoctoral advisor:

1. Pulong Ma (DSS & SAMSI, 01/2019 – 06/2021)
2. Akira Horiguchi (DSS, 01/2021 – 06/2024)
3. Hyotae Kim (DSS, 01/2022 – 05/2024)
4. Ganchao Wei (DSS, 11/2023 –)
5. Yuliang Xu (DSS 07/2024 –)

M.S. thesis advisor:

1. Jialiang Mao (DSS, 11/2015 – 05/2016 degree conferred)
2. Yuhan Chen (DSS, 12/2015 – 05/2016 degree conferred)
3. Hanyu Song (DSS, 08/2016 – 05/2017 degree conferred)
4. Morris Greenberg (DSS, 06/2020 – 05/2021 degree conferred)
5. Jiongran Wang (DSS, 02/2022 – 05/2023 degree conferred)

Undergraduate thesis advisor:

1. Kate Yuan (DSS, graduated in 2012 with honor)
2. Jenny Zhang (DSS, graduated in 2013 with honor)
3. Vivek Sriram (DSS, graduated in 2019 with high honor)

Ph.D. pre-lim and dissertation committee:

1. Lynn Lin (DSS, graduated in 2012)
2. Yun Yang (DSS, graduated in 2014)
3. Douglas VanDerwerken (DSS, graduated in 2015)
4. Timothy Au (DSS, graduated in 2014)
5. Kaoru Irie (DSS, pre-lim in 2014)
6. Zoey Zhao (DSS, graduated in 2015)
7. Brian St. Thomas (DSS, pre-lim in 2013)
8. Yizhe Zhang (CBB, pre-lim in 2015)
9. Christoph Hellmayr (DSS, pre-lim in 2015, graduated in 2017)
10. Sayan Patra (DSS, pre-lim in 2017, , graduated in 2019)
11. Sheng Jiang (DSS, pre-lim in 2018, graduated in 2021)
12. Kimberly Roche (CBB, prelim in 2019, graduated in 2022)
13. Jackie Vahey (CBB, prelim in 2019, graduated in 2022)
14. Xuechan Li (B&B, pre-prelim in 2019, graduated in 2022)
15. Teng Wang (BME, prelim in 2019, graduated in 2023)
16. Yichen Zhu (DSS, prelim in 2020, graduated in 2023)
17. Hanyu Song (DSS, prelim in 2020, graduated in 2022)
18. Benedetta Bruni (DSS, prelim in 2023)
19. Yueqi Guo (DSS, prelim in 2023)
20. Braden Scherting (DSS, prelim in 2023)
21. Hun Kang (DSS, prelim in 2024)
22. Derek Cho (DSS, prelim in 2024)

Ph.D. rotation advisor:

1. Mengqi Zhang (CBB, Summer 2015)
2. Danting Jiang (CBB, Summer 2019)

3. Kuei-Yueh Ko (CBB, Fall 2019)

M.S. thesis committee:

1. Liqi Feng (B&B, graduated in 2015)
2. Brian Cozzi (DSS, graduated in 2020)

Undergraduate thesis committee:

1. Jack Fu (DSS, graduated in 2013)
2. Christine Cheng (DSS, graduated in 2013)
3. Jaslyn Zhang (DSS, graduated in 2017)