Haozhe Zhang

DATA & APPLIED SCIENTIST, PH.D. IN STATISTICS

One Microsoft Way
Redmond, WA 98052
Cell: +1–(515) 708–2685
Email: haozhe.stat@gmail.com
Webpage : haozhestat.github.io
Github : github.com/haozhestat

PROFESSIONAL INTERESTS

Automation, Diagnostics, Inference and Interpretation for Machine Learning, Large-scale Data Analysis, Functional Data Analysis, Online Controlled Experimentation, Applied Data Science.

EDUCATION

Iowa State University, Ames, IA *8/2014 - 8/2019*
Doctor of Philosophy, Statistics
Dissertation: Topics in functional data analysis and machine learning predictive inference
Advisors: Prof. Dan Nettleton and Prof. Yehua Li

University of Science and Technology of China, Hefei, China *8/2010 - 6/2014*
School of the Gifted Young
Hua Loo-Keng Talent Program in Mathematics
Bachelor of Science, Statistics

National Taiwan University, Taipei *Fall 2012*
Department of Mathematics
Exchange Student

PROFESSIONAL EXPERIENCE

Microsoft AI Platform, Bellevue, WA
Data & Applied Scientist 2 *5/2019 - Present*

- Applied Research on automated machine learning algorithms with speed and scale.
- Develop statistical and machine learning models for online controlled experiments.

Laurence H. Baker Center for Bioinformatics and Biological Statistics, Ames, IA
Research Assistant *5/2015 - 05/2019*

- Build predictive models using machine learning algorithms to forecast crop yields in US Midwest.
- Develop methodologies that quantify the uncertainty of predictions from bagging methods (e.g. Random Forests), and develop regression-enhanced random forest to improve forecasting accuracy.
- Analyze genomic, phenotypic, and environmental data from agricultural and biological sciences.

eBay Inc., San Jose, CA *Summer 2018*
Data Scientist Intern

- Develop large-scale graph-based semisupervised learning algorithm utilizing label propagation for look-alike system of eBay ads business.
- Design and perform online experimentations for the proposed graph-based recommendation system and classic SVD-based collaborative filtering.

Department of Statistics, Iowa State University, Ames, IA
Teaching Assistant *8/2017 - 5/2018*

- Lab Instructor of STAT 101 (Principle of Statistics) in the fall semester.
- Lab Instructor of STAT 326 (Introductory Business Statistics II) in the spring semester.

Liberty Mutual, Boston, MA *Summer 2017*
Data Scientist Intern

- Develop and test deep neural networks (CNN, LSTM, ResNet, U-net etc.) for telematics data (i.e., vehicle GPS data) using MXNet and Tensorflow.

- Compare the performance of deep learning with state-of-art machine learning methods such as gradient boosting and explore ensemble opportunities.
- Test and improve Amazon Website Service (AWS) based deep learning platform.

Okinawa Institute of Science and Technology, Okinawa, Japan
Research Intern *Summer 2013, Winter 2012*

- Analyze satellite image data and 3D neuron data using Python (scipy, numpy), and Investigate the spatial-temporal pattern of fairy circle point processes.
- Build a probabilistic model for geometric patterns of neuron bifurcations.

WORKING PAPERS

- **Zhang, H.**, and Li, Y. (2020+). "Unified Principal Component Analysis for Sparse and Dense Functional Data under Spatial Dependency". Manucript in preparation for *Journal of the American Statistical Association*.
- **Zhang, H.**, Nettleton, D., Hey, S., Jubery, Z., Yeh, C.T., and Schnable, P. (2020+). "Estimating Plant Growth Curves and Derivatives by Modeling Crowdsourced Imaged-based Data". Manucript in preparation for *The Annals of Applied Statistics*.
- Wang, C., Wu, Q., Weimer, M., and **Zhang, H.** (2020+). "FLAML: An Fast and Lightweight AutoML Library". Submitted to *NeurIPS 2020*.
- Johnstone, C., **Zhang, H.**, and Nettleton, D. (2020+). "piRF: Implementation of Prediction Interval Methods for Random Forests in R". Manucript in preparation for *Journal of Statistical Software*.

PUBLICATIONS

- Liang, D., **Zhang, H.**, Chang, X., and Huang, H. (2020). "Modeling and Regionalization of China's $PM_{2.5}$ Using Spatial-Functional Mixture Model". *Journal of the American Statistical Association*. DOI: 10.1080/01621459.2020.1764363.
- **Zhang, H.**, Zimmerman, J., Nettleton, D., and Nordman, D.J. (2020). "Random Forest Prediction Intervals". *The American Statistician*. DOI: 10.1080/00031305.2019.1585288.
- Guo, T., Yu, X., Li, X., **Zhang, H.**, Zhu, C., Flint-Garcia, S., McMullen, M., Szalma, S., Holland J., Wisser, R., and Yu, J. (2019). "Optimal Designs for Genomic Selection in Hybrid Crops". *Molecular Plant*, 12(3):390-401. DOI:10.1016/j.molp.2018.12.022.
- **Zhang, H.**, Nettleton, and D., Zhu, Z. (2017). "Regression-Enhanced Random Forests". In JSM Proceedings, Section on Statistical Learning and Data Science, Alexandria, VA: American Statistical Association. 636 – 647.
- **Zhang, H.**, and Sinclair, R. (2015)."Namibian Fairy Circles and Epithelial Cells Share Emergent Geometric Order". *Ecological Complexity*, 22:32-35. DOI:10.1016/j.ecocom.2015.02.001.
- Liang, X., Zou, T., Guo, B., Li, S., **Zhang, H.**, Zhang, S., Huang, H., and Chen, S.X. (2015). "Assessing Beijing's $PM_{2.5}$ Pollution: Severity, Impacts of Weather, APEC and Winter Heating". *Proceedings of the Royal Society A*, 471(2182). DOI: 10.1098/rspa.2015.0257.

DISSERTATION

- **Zhang, H.** (2019). "Topics in Functional Data Analysis and Machine Learning Predictive Inference". *PhD Dissertation. Iowa State University Digital Repository*.17929.

Invited Talks & Posters

- **Zhang, H.**, Qu, Y., and Kirmani, S. "Constructing a Graph from User Implicit Feedback". In: 2019 SIAM Conference on Computational Science and Engineering (CSE19), Spokane, WA, February 2019.
- **Zhang, H.**, Nettleton, D., and Nordman, D.J. "Random Forest Prediction Intervals". In: Data Science and Analytic Team at Facebook, Menlo Park, CA, January 2019.
- **Zhang, H.**, and Li, Y. "Spatially Dependent Functional Data: Covariance Estimation, Principal Component Analysis, and Prediction". In:Department of Mathematics and Statistics, University of New Hampshire, Durham, NH, December 2018.
- **Zhang, H.**, and Li, Y. "Spatially Dependent Functional Data: Covariance Estimation, Principal Component Analysis, and Prediction". In: School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia, December 2018.

Contributed Talks & Posters

- **Zhang, H.**, Nettleton, D., Hey, S., Jubery, Z., Yeh, C.T., and Schnable, P. "Estimating Plant Growth Curves and Derivatives by Modeling Crowdsourced Imaged-based Data". In: Symposium on Data Science and Statistics, Bellevue, WA, June 2019.
- **Zhang, H.**, and Li, Y. "On the Covariance Estimation and Principal Component Analysis for Spatially Dependent Functional Data". In: Joint Statistical Meetings, Vancouver, Canada, August 2018.
- **Zhang, H.**, Nettleton, D., and Nordman, D.J. "Random Forest Prediction Intervals". In: Symposium on Data Science and Statistics, Reston, Virgina, May 2018.
- **Zhang, H.**, and Li, Y. "On the Covariance Estimation and Principal Component Analysis for Spatially Dependent Functional Data". In: ENAR Spring Meeting, Atlanta, Georgia, March 2018.
- **Zhang, H.**, Nettleton, D., and Zhu, Z. (2017). "Regression-Enhanced Random Forests". In: Joint Statistical Meetings, Baltimore, Maryland, August 2017.
- **Zhang, H.**, and Nettleton, D. "Prediction-Guided Statistical Ranking of Maize Seed Brands". In: Symposium on Predictive Crop Design: Genome-to-Phenome, Lincoln, Nebraska, April 2017.
- **Zhang, H.**, and Zhu, Z. "Clustering Multiscale Spatial Functional Data with Application to Precipitation Regimes Identification". In: the 6th International Workshop on Climate Informatics, Boulder, Colorado, September 2016.
- **Zhang, H.**, and Zhu, Z. "Clustering Multiscale Spatial Functional Data with Application to Precipitation Regimes Identification". In: Workshop on Bayesian environmetrics, Columbus, Ohio, March 2016.

Selected Awards & Honors

- **Student & Early Career Travel Award**, 2019. Travel fund for presenting an e-poster in the 2019 Symposium on Data Science and Statistics in Bellevue, Washington.
- NSF Travel Grant for attending the ACM-IMS Interdisciplinary Summit on the Foundations of Data Science inSan Francisco, CA on June 2019.
- **SAMSI Travel Award**, 2018. Travel fund for presenting a paper in the Symposium on Data Science and Statistics in Reston, Virginia.
- Travel support from ASA Section for Statistical Programmers and Analysts, 2017-18.
- **The George W. Snedecor Award**, Department of Statistics, Iowa State University, 2016. This award is for the most outstanding Ph.D. candidate in the department.

- Climate Informatics Travel Fellowship Award, 2016. Travel fund for giving a spotlight oral presentation in the 6th International Workshop on Climate Informatics, Boulder, CO.
- **The Holly C. and E. Beth Fryer Award**, Department of Statistics, Iowa State University, 2015. This award is for a top second-year Ph.D. student in the department.
- **2nd Place** and 5th Place at Data Mining Cup 2016, Prudsys AG, Berlin, Germany.
- **Presidential Scholars Fellowship**, Iowa State University, 2014.
- **National Scholarship of China**, Ministry of Education of the People's Republic of China, 2013. This award is for outstanding undergraduate students in China.
- Outstanding Student Scholarship, School of the Gifted Young, University of Science and Technology of China, 2012.
- National Encouragement Scholarship, School of the Gifted Young, University of Science and Technology of China, 2011.

Programming Skills

- **Projects in**: Python, R, SQL, C/C++, Spark, AzureML.
- **Familiar with**: Unix shell, HTML.
- **IDEs**: vi/vim,Visual Studio, PyCharm, Jupiter.

Service and Leadership

- Proposal Reviewer, *Fall 2019 MLADS Conference* (Microsoft machinel learning internal conference).
- Session Chair, *Spatial and Spatiotemporal Modeling in Climate and Meteorology*, Section on Statistics and the Environment, Joint Statistical Meetings 2019.
- Program Committee Member, Artificial Intelligence Section, 2019 Grace Hopper Celebration of Women in Computing.
- Reviewer for *Statistica Sinicia* (1), *KDD'19* (2), *IEEE Transactions on Knowledge and Data Engineering* (1), *Stat* (1), *PLOS one* (1) and *PeerJ* (3).
- International Student Advisory Board, Iowa State University, 2018.
- Computation Advisory Committe, Department of Statistics, Iowa State University, 2017.
- President (2017–18) and Treasurer (2016–17) of the Iowa STAT-ers, 2017–18.
- Graduate and Professional Student Senate, Iowa State University, 2015–16.
- Student Assistant, Office of International Affairs, University of Science and Technology of China, 2013–14.