

# Haozhe Zhang

DATA & APPLIED SCIENTIST, PH.D. IN STATISTICS

3219 Snedecor Hall  
2438 Osborn Dr.  
Ames, IA 50011, U.S.A.  
Cell: +1-(515) 708-2685  
Email: [haozhe.stat@gmail.com](mailto:haozhe.stat@gmail.com)  
Webpage : [haozhestat.github.io](http://haozhestat.github.io)  
Github : [github.com/haozhestat](https://github.com/haozhestat)

---

PROFESSIONAL INTERESTS	Large-scale Data Analysis, Diagnostics and Uncertainty Quantification of Machine Learning Algorithms, Functional Data Analysis, Predictive Analytics, Design of Online Experiments, and Interdisciplinary Quantitative Research.
------------------------	--

---

EDUCATION	<b>Iowa State University, Ames, IA</b> <i>Doctor of Philosophy, Statistics</i> Advisors: Prof. Dan Nettleton and Prof. Yehua Li 8/2014 - 5/2019
	<b>Iowa State University, Ames, IA</b> <i>Master of Science, Statistics</i> Advisor: Prof. Zhengyuan Zhu 8/2014 - 5/2016
	<b>University of Science and Technology of China, Hefei, China</b> School of the Gifted Young Hua Loo-Keng Talent Program in Mathematics <i>Bachelor of Science, Statistics</i> 8/2010 - 6/2014
	<b>National Taiwan University, Taipei</b> Department of Mathematics <i>Exchange Student</i> Fall 2012

---

PROFESSIONAL EXPERIENCE	<b>Laurence H. Baker Center for Bioinformatics and Biological Statistics, Ames, IA</b> <i>Research Assistant</i> 5/2015 - 05/2019 <ul style="list-style-type: none"><li>- Build predictive models using machine learning algorithms to forecast crop yields in US Midwest.</li><li>- 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.</li><li>- Analyze genomic, phenotypic, and environmental data from agricultural and biological sciences.</li></ul>
	<b>eBay Inc., San Jose, CA</b> <i>Data Scientist Intern</i> Summer 2018 <ul style="list-style-type: none"><li>- Develop large-scale graph-based semisupervised learning algorithm utilizing label propagation for look-alike system of eBay ads business.</li><li>- Design and perform back test and online A/B test for the proposed graph-based recommendation system and classic SVD-based collaborative filtering.</li></ul>
	<b>Department of Statistics, Iowa State University, Ames, IA</b> <i>Teaching Assistant</i> 8/2017 - 5/2018 <ul style="list-style-type: none"><li>- Lab Instructor of STAT 101 (Principle of Statistics) in the fall semester.</li><li>- Lab Instructor of STAT 326 (Introductory Business Statistics II) in the spring semester.</li></ul>
	<b>Liberty Mutual, Boston, MA</b> <i>Data Scientist Intern</i> Summer 2017 <ul style="list-style-type: none"><li>- Develop and test deep neural networks (CNN, LSTM, ResNet, U-net etc.) for telematics data (i.e., vehicle GPS data) using MXNet and Tensorflow.</li></ul>

- 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.

**Center for Survey Statistics and Methodology, Ames, IA**

*Research Assistant*

8/2014 - 6/2016

- Work on interdisciplinary survey sampling projects, including National Resources Inventory (NRI) and Conservation Effects Assessment (CEA) projects.
- Use image-to-image neural network to predict landuse categories on remote sensing data.

**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.

---

MANUSCRIPTS  
SUBMITTED

- **Zhang, H.**, and Li, Y. (2019+). "Spatially Dependent Functional Data: Covariance Estimation, Principal Component Analysis, and Kriging". Submitted to *The Annals of Statistics*.
- Liang, D., **Zhang, H.**, Chang, X., and Huang, H. (2019+). "Modeling and Regionalization of China's PM<sub>2.5</sub> Using Spatial-Functional Mixture Model". Submitted to *Journal of the American Statistical Association*.

PUBLICATIONS

- **Zhang, H.**, Zimmerman, J., Nettleton, D., and Nordman, D.J. (2019+). "Random Forest Prediction Intervals". Accepted by *The American Statistician*.
- Guo, T., Yu, X., Li, X., **Zhang, H.**, Zhu, C., Flint-Garcia, S., McMullen, M., Szalma, S., Holland J., Wissner, 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, D., and 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.**, Zhu, Z., and Yin, S. (2016). "Identifying Precipitation Regimes in China Using Model-based Clustering of Spatial Functional data". *Proceedings of the Sixth International Workshop on Climate Informatics*, pages 117–120. DOI:10.5065/d6k072n6.
- He, Z., Zhang, M., and **Zhang, H.** (2016). "Data-driven Research on Chemical Features of Jingdezhen and Longquan Celadon by Energy Dispersive X-ray Fluorescence". *Ceramics International*, 42(4):5123–5129. DOI: 10.1016/j.ceramint.2015.12.030.
- Liang, X., Zou, T., Guo, B., Li, S., **Zhang, H.**, Zhang, S., Huang, H., and Chen, S.X. (2015). "Assessing Beijing's PM<sub>2.5</sub> Pollution: Severity, Impacts of Weather, APEC and Winter Heating". *Proceedings of the Royal Society A*, 471(2182). DOI: 10.1098/rspa.2015.0257.
- **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). "Air Quality Assessment Report I: Statistical Analysis of Beijing". *Technical Report (in Chinese)*, 2015.

---

## WORKING PAPERS

- "Estimating Plant Growth Curves and Derivatives by Modeling Crowdsourced Imaged-based Data". Joint work with Dan Nettleton, Pat Schnable, Stefan Hey, and Zaki Jubery. Manuscript in preparation.
- "Constructing a Graph from User Implicit Feedback". Joint work with Shad Kirmani and Yan Qu. Manuscript in preparation.
- "Neonate Garter Snakes *Thamnophis Elegans* Exhibit Consistent Among-individual Variation in Behavior and Habituation at Multiple Time Scales". Joint work with Eric J. Gangloff, Vianey Leos, and Anne Bronikowski. Manuscript in preparation for *Animal Behavior*.

---

## INVITED TALKS & POSTERS

\* Denotes presenter.

- **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.**, 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, Virginia, 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

- **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++, Hive, Hadoop, Matlab, LATEX.
- **Familiar with:** Unix shell, SAS, HTML.
- **IDEs:** vi/vim, Visual Studio, PyCharm, Jupiter.

---

## SERVICE AND LEADERSHIP

- 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 Sinica* (1), *Stat* (1), and *PeerJ* (3).
  - International Student Advisory Board, Iowa State University, 2018.
  - Computation Advisory Committee, Department of Statistics, Iowa State University, 2017.
  - President of the Iowa STAT-ers, 2017–18.
  - Treasurer of the Iowa STAT-ers, 2016–17.
  - 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.
-

PROFESSIONAL  
AFFILIATION

- American Statistical Association (ASA), Member, 2014 - Present
  - Association for Computing Machinery (ACM), Member, 2019 - Present
  - Institute of Mathematical Statistics (IMS), Member, 2014 - Present
  - International Chinese Statistical Association (ICSA), Member, 2015 - Present
  - Society for Industrial and Applied Mathematics (SIAM), Member, 2017 - Present
- 

REFERENCE

Dan Nettleton, Ph.D.  
Distinguished Professor  
Department of Statistics  
Iowa State University  
Phone: +1-(515)-294-7754  
Email: [dnett@iastate.edu](mailto:dnett@iastate.edu)

Yehua Li, Ph.D.  
Professor  
Department of Statistics  
University of California at Riverside  
Phone: +1-(951)-827-3327  
Email: [yehuali@ucr.edu](mailto:yehuali@ucr.edu)

Song Xi Chen, Ph.D.  
University Chair Professor  
Guanghua School of Management  
Peking University  
Phone: +86-10-62750427  
Email: [csx@gsm.pku.edu.cn](mailto:csx@gsm.pku.edu.cn)