

# Yu Michael Zhu

Address: Department of Statistics, Purdue University  
West Lafayette, IN 47907-1399  
Telephone: (765) 494-6038  
Fax: (765) 494-0558  
E-mail: yuzhu@purdue.edu  
WWW: <http://www.stat.purdue.edu/~yuzhu>

## RESEARCH INTERESTS

Generative Modeling of Text and Image Data  
Dimension Reduction and Variable Selection for High Dimensional Data Analysis  
Design and Analysis of Experiments for Industrial and Medical Applications

## EDUCATION

Ph.D. in Statistics (2000) University of Michigan, Ann Arbor, MI  
M.S. in Applied Mathematics (1994) Tsinghua University, Beijing, China  
B.S. in Applied Mathematics (1990) Tsinghua University, Beijing, China

## PROFESSIONAL EXPERIENCE

Purdue University	Professor	2013 – Present	W. Lafayette, IN
Tsinghua University	Visiting Professor	2016–2018	Beijing, China
Purdue University	Associate Professor	2006–2013	W. Lafayette, IN
Purdue University	Assistant Professor	2000–2006	W. Lafayette, IN

## PUBLICATION

### Published and Submitted

1. Yang, Y., Deng, K. and Zhu, M. (2022). Multi-Level Training and Bayesian Optimization for Hyper-Parameter Optimization. Submitted to Technometrics.
2. Reese, I. and Zhu, M.Y. (2022). Hierarchically Decodable Output Codes for Transparent Large Scale Image Classification. Submitted.
3. Luo, Z.X. and Zhu, M.Y. (2022). Recurrent Neural Networks with Deep Hierarchical Mixed Structures for Chinese Document Classification. Submitted.
4. Hanmin Guo. Lin Hou. Yu Zhu. (2022) Minimal  $\phi$ -field for flexible sufficient dimension reduction. *Electron. J. Statist.* 16 (1) 1997 -2032. <https://doi.org/10.1214/22-EJS1999>
5. Zhaoxin Luo and Michael Zhu. (2022). Recurrent Neural Networks with Mixed Hierarchical Structures and EM Algorithm for Natural Language Processing. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 61046113, Marseille, France. European Language Resources Association.
6. Xin Hao and Michael Zhu. (2022). Score-based Image-to-Image Regression with Joint Diffusion. In *Proceedings of the 21st International Conference on Machine Learning and Applications*, Bahamas.

7. Reese, T., and Zhu, Y.M. (2021 online). LB-CNN: Convolutional Neural Network with Latent Binarization for Large Scale Multi-class Classification. In: Wani, M.A., Raj, B., Luo, F., Dou, D. (eds) Deep Learning Applications, Volume 3. *Advances in Intelligent Systems and Computing*, vol 1395. Springer, Singapore.
8. Luo, Z. and Zhu, M. (2021). Recurrent Neural Networks with Mixed Hierarchical Structures for Natural Language Processing, accepted by IJCNN 2021: The International joint conference on neural networks.
9. Xin, H. and Zhu, Y.M. (2020). Conditional Score Matching for Image to Image Regression. In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2020, pp. 407-414, DOI: 10.1109/ICMLA51294.2020.00071.
10. Reese T. and Zhu M. (2020). LB-CNN: Convolutional Neural Network with Latent Binarization for Large Scale Multi-class Classification. In: 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), pp142-147, DOI: 10.1109/ICMLA51294.2020.00031.
11. Cho, H., Yu, B., Cannon, J. and Zhu, Y.M. (2018). Efficacy of a Media Literacy Intervention for Indoor Tanning Prevention. *Journal of Health Communication*. To Appear.
12. Liu, S., Chen, J., Sun, Z. and Zhu, M. (2018). From Good to Great: Nonlinear Improvement of Healthcare Service. *International Journal of Pharmaceutical and Healthcare Marketing*. To Appear.
13. Zeng, P. and Zhu, Y. (2018). A Novel Regularization Method for Estimation and Variable Selection in Multiple Index Models. *Communications in Statistics–Theory and Methods*. To Appear
14. Zhang, R. Deng, W. Zhu, Y. (2017). Using Deep Neural Networks to Automate Large Scale Statistical Analysis for Big Data Applications. *Proceedings of the 9th Asian Conference on Machine Learning (ACML17), Seoul, Korea, 2017*.
15. Pan, C. and Zhu, M (2017). Group Additive Structure Identification for Kernel Non-parametric Regression. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*
16. Huang, Q. and Zhu, M. (2017). SPOT: Sparse Optimal Transformations for High Dimensional Variable Selection and Exploratory Regression Analysis. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2017)*
17. Cheng, L, Zeng, P, and Zhu, Y. (2017). BS-SIM: An Effective Variable Selection Method for High-dimensional Single Index Model. *Electronic Journal of Statistics*. Vol 11, 2 , 3522-3548.
18. Zhang, R., Hu, M., Zhu, Y., Qin, S. Q, Deng, K. and Liu, J.S. (2017). Inferring Spatial Organizations of Individual Topologically Associated Domains via Chromosomes via

piecewise Helical Models. *IEEE Transactions on Computational Biology and Bioinformatics* .

19. Li, B., Sun, Z., He, Q. Zhu, Y. and Qin, Z. (2016). Bayesian inference with historical data- based informative priors improves detection of differentially expressed genes. *Bioinformatics*. 32, 682–689.
20. Wu, H. and Zhu, Y. (2015). Deconvolution of Base-Level RNA-Seq Read Count for Quantification of Transcript Expression Levels. *The Annals of Applied Statistics*. 10, 1195–1216.
21. Huang, Q. and Zhu, Y. (2016). Model-Free Sure Screening Via Maximum Correlation. *Journal of Multivariate Analysis*. 148, 89–106
22. Pan, C., Huang, Q., and Zhu, M. (2015). Optimal Kernel Group Transformation for Exploratory Regression Analysis and Graphics. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp. 905-914. ACM.
23. Wei, W., Liu, C, Zhu, M.Y. and Matei, S. (2015). A Nonparametric Hidden Markov Clustering Model with Applications to Time Varying User Activity Analysis. In *Proceedings of 14th IEEE International Conference on Machine Learning and Applications (ICMLA 2015)*.
24. Matei, S.A., Bertino, E., Zhu, M., Liu, C., Si, L., and Britt, B. (2015). A Research Agenda for the Study of Entropic Social Structural Evolution, Functional Roles, Adhocratic Leadership Styles, and Credibility in Online Organizations and Knowledge Markets. In *Roles, Trust, and Reputation in Social Media Knowledge Markets*. pp333, Springer.
25. Chen, L. and Zhu, Y. (2014). A classification approach for DNA methylation profiling with bisulite next-generation sequencing data. *Bioinformatics*. 15, 172–179.
26. Sun, Z. Kuczek, T, and Zhu, Y. (2014) Statistical Calibration of qRT-PCR, Microarray, and RNA-Seq Gene Expression Data with Measurement Error Models. *Annals of Applied Statistics*. 8, 1022–1044.
27. Kwon, B.C., Yi, J.S. and Zhu, Y. (2013). ReadingMate The Effect of the Content Stabilizing Technique, Font Size, and Interline Spacing on the Letter-Counting Task Performance of Treadmill Runners. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 1101-1111.
28. Wan. H., Zhu. Y., Sanchez. S. M., Xing, D. and K. Turgut. (2013) Simulation screening experiments using lasso-optimal supersaturated design and analysis: A maritime operations application. In *Proceedings of the 2013 Winter Simulation Conference (WSC)* edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 497508.
29. Wu, H., Qin, Z., and Zhu, Y. (2013). MP-Seq: Using Finite Poisson Mixture Models for RNA-Seq Data Analysis and Transcript Expression Level Quantification. *Statistics in Biosciences*.

30. Guo Y., Zhu Y., Salvendy G., Proctor R.W. (2013) Value Added by the Axiomatic Usability Method for Evaluating Consumer Electronics. In *Human Interface and the Management of Information. Information and Interaction for Learning, Culture, Collaboration and Business*. HIMI 2013. Lecture Notes in Computer Science, vol 8018, 457–466. Springer.
31. Zhong, W., Zhang, T., Zhu, Y and Liu, J. S. (2012). Correlation Pursuit: Forward Stepwise Variable Selection for Index Models. *Journal of Royal Statistical Society, Ser B*. 74, 849–870.
32. Sun, Z. and Zhu, Y. (2012). Systematic Comparison of RNA-Seq Normalization Methods Using System of Measurement Error Models. *Bioinformatics*. 28:20. 2584–2591.
33. Hu M, Zhu Y, Taylor JMG, Liu JS, Qin ZS. (2012). Using Poisson Mixed-Effects Model to Quantify Exon-Level Gene Expression In RNA-seq. *Bioinformatics*. 28, 63-68
34. Zeng, P., He T. and Zhu, Y. (2012). A Lasso-Type Approach for Estimation and variable Selection in Single Index Models *Journal of Computational and Graphical Statistics*. 21, 92-109.
35. Zhang, T., Wan, H. and Zhu, Y. (2012). Detection and Localization of Hidden Radioactive Sources with Spatial Statistical Methods. *Annals of Operation Research*. 192, 87-104.
36. Zeng, P., Wan, H., and Zhu, Y. (2011). A Coset Patter Identity Between a  $2^{n-p}$  Design and its Complement. *Statistica Sinica*. 12, 1453-1471.
37. Tokdar, S., Zhu, Y., and Ghosh, J. K. (2010). Bayesian Density Regression with Logistic Gaussian Process and Subspace Projection. *Bayesian Analysis*, 5, 319-344.
38. Zeng, P. and Zhu, Y. (2010). An Integral Transform Method for Estimating Central Mean and Central Subspaces. *Journal of Multivariate Analysis*, 101, 271-290.
39. Zeng, P., Zhu, Y. and Phelps, K. T. (2009). Bounds on the Maximum Number of Factors in Designs with Two Distinct Groups of Factors. *Communications in Statistics-Theory and Methods* .39, 2302–2310.
40. Liu, L., Levine, M. and Zhu, Y. (2009). A Functional EM Algorithm for Mixing Density Estimation via Nonparametric Penalized Likelihood Maximization. *Journal of Computational and Graphical Statistics*. Vol 18, 481–504.
41. Fan, X., Liu, S. S., and Zhu, Y. (2008). An Innovative Approach Examining the Asymmetrical and Nonlinear Relationship between Attribute-Level Performance and Service Outcomes, *Advances in consumer Research*, 36, 569-570.
42. Zhu, Y., Zeng, P. and Jennings, K. (2007). Optimal Compound Orthogonal Arrays and Single Arrays for Robust Parameter Design Experiments. *Technometrics*. Vol 49, 440–453.

43. Zhu, Y. and Wu, C.F.J. (2007). Structure Functions for Regular  $s^{l-m}$  Designs With Multiple Groups of Factors. *Statistica Sinica*. Vol. 17, 1239–1259.
44. Liu, L. and Zhu, Y. (2007). Partially Projected Gradient Algorithms for Calculating Non-parametric Maximum Likelihood Estimates of Mixing Distributions. *Journal of Statistical Planning and Inferences*. Vol. 137, 2509–2522.
45. Zhu, Y. and Zeng P. (2006). Fourier Methods for Estimating the Central Subspace and the Central Mean Subspace in Regression. *Journal of the American Statistical Association*. Vol. 101, 1638–1651.
46. Feng, R., Cai, L., and Zhu, Y. (2006). Long Haul Travel Attitude Construct and Relationship to Behavior—A Case of French Travelers. *Journal of Travel and Tourism Marketing*. Vol. 20, 1-11.
47. Zhong, W., Peng, Z., Ma, P., Liu, J.S. and Zhu, Y. (2005). RSIR: Regularized Sliced Inverse Regression for Motif Discovery. *Bioinformatics*. Vol. 21, 4169–4175.
48. Zhu, Y. and Zeng, P. (2005). On the Coset Pattern Matrices and Minimum  $M$ -Aberration of  $2^{n-p}$  Designs. *Statistica Sinica*. Vol.15, 717-730.
49. Lin, X., Clifton, C., and Zhu, M. (2005). Privacy Preserving Clustering with Distributed EM Mixture Modeling. *Knowledge and Information Systems, An International Journal*. Vol.8, No.1, 68-81.
50. Zhu, Y. and Liu, L. (2004). Optimal Randomization for Privacy Preserving Data Mining. In *KDD '04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 761-766. (25.2%)
51. Lin, X. and Zhu, Y. (2004). A Degenerate Expectation-Maximization Algorithm for Local Dimension Reduction. In *Proceedings of the 2004 meeting of International Federation of Classification Society*. 259-269. (78%)
52. Zhu, Y. (2003). Structure Function for Aliasing Patterns in  $2^{l-n}$  Design with Multiple Groups of Factors. *The Annals of Statistics*. Vol. 31, No. 3, 995-1011.
53. Wu, C.F.J. and Zhu, Y. (2003). Optimal Selection of Single Arrays for Parameter Design Experiments. *Statistica Sinica*. Vol. 12, No. 4, 1179-1199.
54. Clifton, C., Kantarcioglu, M., Lin, X., Vaidya, J. and Zhu, Y. (2002). Tools for Privacy Preserving Distributed Data Mining. *SIGKDD Explorations*. Vol.4, No.2, 28-34.

## Book Chapters and Books

1. Reese T. and Zhu M. (2021). LB-CNN: Convolutional Neural Network with Latent Binarization for Large Scale Multi-class Classification (Extended Version). To appear as a book chapter in *Deep Learning Applications Volume 3*, edited by Wani, M. A.

2. Sun, Z., Wu, H., Qin, Z., and Zhu, Y. (2013). *Model-Based Methods for Transcript Expression Level Quantification in RNA-Seq* in Advances in Statistical Bioinformatics: Models and Integrative Inference for High-Throughput Data *edited by* : Do, K-A., Qin, S. and Vannucci, M. *Cambridge University Press*.
3. Vaidya, J., Clifton, C. and Zhu, M. (2005). *Privacy Preserving Data Mining*. Advances in Information Security Vol.19. *Springer-Verlag*.
4. Zhu, Y. (2008). *Product Array Designs* in Encyclopedia of Statistics in Quality and Reliability *edited by* Ruggeri, F, Kenett, R. S. and Ialtis, F. W. *Wiley*

## INVITED LECTURES

1. The 21st International Conference on Machine Learning and Applications, Bahamas, December, 2022. *Score-based Image-to-Image Regression with Joint Diffusion*.
2. Virtual Talk: Conditional Score Matching for Image to Image Regression, April 16, 2021, Bear Club Organization, Beijing, China
3. The 10th ICASA International Conference on Global Growth of Modern Statistics in the 21th Century, December, Shanghai, China, December, 2016. *SPOT: Sparse Optimal Transformations for High Dimensional Variable Selection and Exploratory Regression Analysis*.
4. International Workshop on Brain Health and Important Diseases, South China University of Technology, GuangZhou, June, 2016, *Advances in Statistical Methodology for Neuroimaging Data Analysis*
5. Center for Statistical Science, Tsinghua University, Beijing, China, March, 2016: *Optimal Kernel transformation for exploratory regression analysis and graphics*
6. Lecture Series on the Frontier of Big Data Technology, The Institute of Big Data Science, Tsinghua University, Beijing: *Applications and Perspectives of the R Programming Language in Big Data Statistical Computing and Analysis*
7. Spring Research Conference, Cincinnati, OH, USA, 2015. *Lasso and Supersaturated Design: Analysis, Optimality Criteria, and Algorithms*
8. School of Science, Beijing Technology and Business University, Beijing China, October, 2013, *Quantification of Gene and Transcript Expression Levels Using RNA-Seq Data*.
9. Department of Statistics, Peking University, Beijing, China, October, 2013, *Quantification of Gene and Transcript Expression Levels Using RNA-Seq Data*.
10. Department of Mathematics/Mathematical Science Center, Tsinghua University, Beijing, China, December, 2013, *Statistical Calibration of Gene Expression Data Using Measurement Error Models*.

11. JSM, San Diego, CA, August 2012, *Using Finite Poisson Mixture Models for RNA-Seq Data Anylysis and Transcript Expression Level Quantification.*
12. ICSA Applied Statistics Symposium, Boston, June 2012, *Systematic Comparison of RNA-Seq Normalization Methods Using Measurement Error Models*
13. ENAR, Washington DC, April 2012, *Using Finite Poisson Mixture Models for RNA-Seq Data Analysis and Transcript Expression Level Quantification.* (Did not go due to illness and presented by co-author Steve Qin)
14. Department of Statistics, Purdue University , January, 2012, *Using Finite Poisson Mixture Models for RNA-Seq Data Analysis and Transcript Expression Level Quantification.*
15. Center of Applied Statistics, East Normal University of China, Shanghai, 2011 *Correlation Pursuit: Stepwise Variable Selection for High Dimensional Semiparametric Regression.*
16. Department of Mathematics, Tsinghua University, Beijing, China, June, 2011 *Functional EN Algorithm for Penalized Maximum Nonparametric Likelihood Estimation of Mixing Density Functions*
17. Center for Quality Science, Academy of Mathematics and System Science, Beijing, China, June, 2011 *Constructing Optimal Supersaturated Design via Lasso*
18. Department of Statistics, Beijing University, Beijing, China, June, 2011 *Correlation Pursuit: Stepwise Variable Selection for High Dimensional Semiparametric Regression.*
19. Department of Mathematics, Zhejiang University, Hangzhou, China, June, 2011 *Correlation Pursuit: Stepwise Variable Selection for High Dimensional Semiparametric Regression.*
20. IISA 2011 Meeting, Raleigh, NC, April, 2011, *Bayesian Methods for Gene Expression quantification in RNA-Seq*
21. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, March, 2011, *Correlation Pursuit: Stepwise Variable Selection for High Dimensional Semiparametric Regression.*
22. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, November, 2010, *Correlation Pursuit: Stepwise variable Selection for High Dimensional Semiparametric Regression.*
23. Joint Research Conference on Statistics in Quality, Industry and Technology, May 25-May 27, 2010, NIST, Gaithersburg, MD, *Optimal Supersaturated Design for Variable Selection via Lasso.*
24. Department of Statistics, University of Wisconsin, Madison, April 15, 2010, *Correlation Pursuit: Stepwise Variable Selection for Index Models.*
25. Department of Statistics, University of Illinois at Urbana-Champaign, Feb. 18, 2010, *Some Results on the Complementary Approach for Fractional Factorial Design*

26. The Classification Society and Interface Society Annual 2009 Meetings, Washington University School of Medicine, St. Louis, MO, 2009 *Sparse Covariance Thresholding for High Dimensional Regression and Classification*
27. Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, Jun, 2009, *Sparse Covariance Thresholding for High Dimensional variable Selection*
28. Spring Research Conference on Statistics in Industry and Technology, University of Simon Fraser, Vancouver, Canada, May, 2009, *Sparse Covariance Thresholding for High Dimensional Variable Selection.*
29. Spring Research Conference on Statistics in Industry and Technology, School of Industrial and Systems Engineering, Atlanta, GA, May, 2008, *Penalized Linear Methods for Estimation and Variable Selection in Index Models.*
30. Academy of Mathematics and Systems Science, P. R. China, June, 2007, *A Fourier Method for Sufficient Dimension Reduction in Regression*
31. Department of Mathematical Science, Tsinghua University, P. R. China, June, 2007, *An Integral Transform Method for Estimating the Central Mean and Central Subspaces.*
32. Design and Analysis of Experiments Conference (DAE) 2005, Santa Fe, New Mexico, *RSIR: Regularized Slice Inverse Regression for Motif Discovery.*
33. Joint Statistical Meeting, Minneapolis, 2005, *Fourier Methods for Estimating Dimension Reduction Subspaces in High Dimensional Regression without Distributional Constraints.*
34. 2005 Quality and Productivity Research Conference, Minneapolis, Minnesota, 2005, *Optimal Compound Orthogonal Array and Single Array for Robust Parameter Design Experiments.*
35. International Conference on Design of Experiments: Theory and Applications, Memphis, 2005, *Coset Pattern and Minimum M-Aberration Criterion.*
36. Statistics Department, University of Chicago, 2005, *Fourier Methods for Estimating Central Subspace in High-dimensional Regression.*
37. International Conference on Statistics, Combinatorics and Related Areas, 2003, Portland, Maine. *Structure functions for Aliasing Patterns in  $s^{l-m}$  Designs and Regular  $(s^{l-r}, s^r)$  Blocked Designs and Their Complementary Designs.*
38. Statistics Department, Purdue University, 2003. *General Structure Functions of  $s^{l-m}$  Designs.*
39. School of Science, Purdue University, 2002, *Treasure Hunting: A Reviving Profession? (Statistical Data Mining in Massive Data).*
40. Vigre Seminar, Department of Statistics, Purdue University, 2002, *Orthogonal Arrays and Design of Experiment.*



41. Department of Information and Decision Sciences, University of Illinois at Chicago, 2002, *A General Theory for Experimental Design with Multiple Groups of Factors*.
42. Mathematics Department, Western Washington University, 2001, *Optimal Selection of Single Arrays for Robust Parameter Design Experiments*.
43. Statistics Department, University of Illinois at Urbana-Champaign, 2001, *Genetic Regulatory Networks Reconstruction: A Tutorial*.
44. Joint Research Conference on Statistics in Quality, Industry and Technology, 2000, Seattle, WA, *Optimal Planning Techniques for Parameter Design Experiments*. (joint with C. F. Jeff Wu).
45. Joint Statistical Meeting, 1999, Baltimore, MD, *Optimal Selection of Single Arrays for Parameter Design Experiments*.
46. Biometrics Department at Parke-Davis, 1999, Ann Arbor, MI, *Structure-Property-Bioactivity Correlation in Drug Discovery Research*.
47. Operation Technology Center at AT&T, 1998, Middletown, NJ, *Statistical Performance Evaluation for Microsoft IIS 3.0*.