

Tao Shi

Department of Statistics
Department of Computer Science and Engineering
The Ohio State University
317 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210
Tel: (614) 292-4516 Fax: (614) 292-2096
Email: taoshi@stat.osu.edu Web: www.stat.osu.edu/~taoshi

Education

Ph.D. in Statistics, University of California at Berkeley, 2005 (Advisor: Professor Bin Yu)
Polar Cloud Detection using Satellite Data with Analysis and Application of Kernel Learning Algorithms

M.S. in Statistics, Peking University, Beijing, China, 2000

B.S. in Statistics, Peking University, Beijing, China, 1997

Research Interests

Statistical methodology and computation on massive datasets
Statistical properties of machine learning algorithms
Applications of statistics in environmental, atmospheric and geosciences

Professional Experience

Assistant Professor (courtesy appointment) Winter 2009 -
Department of Computer Science and Engineering, The Ohio State University

Assistant Professor Autumn 2005 -
Department of Statistics, The Ohio State University

Visiting Scientist Summer, 2005
National Center for Atmospheric Research (visiting Dr. Doug Nychka)

Graduate Student Researcher 2001-2005
Department of Statistics, University of California, Berkeley (supervisor: Prof. Bin Yu)

Statistics Consultant Spring 2004 and Fall 2001
Statistical Consulting Service, U.C. Berkeley (supervisors: Prof. David Freedman and Prof. Bin Yu)

Teaching Experience

Course Instructor, Department of Statistics, The Ohio State University Fall 2005 -
STAT 528, 529, 656, 693, 723, 760, 882, 893, and 999

Graduate Student Instructor, Department of Statistics, U.C. Berkeley 2001-2005

Course Instructor, Peking University, China 1998

Grant Supports

- **National Science Foundation** 2009 - 2010
“Support for Conference on Nonparametrics Statistics and Statistical Learning (2010)”
Co-investigator (with Joseph Verducci as PI)

- **OSU Dentistry CCTS Pilot Grant** 2009 - 2010
“Intervention Strategies to Improve Periodontal Surgical Outcomes in Smokers”
Co-investigator (with Dimitris N. Tataki as PI)
- **National Aeronautics and Space Administration** 2006 - 2009
“A Comprehensive Statistical Analysis System to Associate Local Land-Cover/Land-Use change and Regional Aerosol Composition and Concentration”
Co-PI (with Darla Munroe as PI, Kate Calder and Ningchuan Xiao as Co-PIs)
- **The Ohio Supercomputer Center** 2006 -
“Super Computer Support for A Comprehensive Statistical Analysis System to Associate Local Land-Cover/Land-Use change and Regional Aerosol Composition and Concentration”
Principle Investigator (with Kate Calder, Darla Munroe, and Ningchuan Xiao as Co-PIs)
- **The Ohio Supercomputer Center** 2005 -
“Super Computer Support for Statistical Analysis on Massive Datasets”
Principle Investigator

Awards

NSF Travel Award to the 7 th World Congress of Probability and Statistics	2008
Travel Award to Interface 2006	2006
Travel Award to Graybill Conference	2005
Outstanding Graduate Student Instructor, U.C. Berkeley	2004

Publications and Manuscripts

Journal Articles

- Data Spectroscopy: Eigenspaces of Convolution Operators and Clustering
Shi, T., Belkin, M., and Yu, B. (2009) *Annals of Statistics*, forthcoming.
- Spatial Characteristics of the Difference between MISR and MODIS Aerosol Optical Depth Retrievals over Mainland Southeast Asia
Xiao, N., **Shi, T.**, Calder, C. A., Munroe, D. K., Berrett, C. and Wolfinbarger, S., and Li, D. (2009) *Remote Sensing of Environment* **113**, 1-9.
- The Relationships Between Biomass Burning, Land-cover/use Change, and the Distribution of Carbonaceous Aerosols in Mainland Southeast Asia: A Review and Synthesis
Munroe, D. K., Wolfinbarger, S.R., Calder, C. A., **Shi, T.**, Xiao, N., Lam, C.Q. and Li, D. (2008) *Journal of Land Use Science* **3**, 161-183.
- Daytime Arctic Clouds Detection Based on Multi-angle Satellite Data with Case Studies
Shi, T., Yu, B., Clothiaux, E. E., and Braverman, A. J. (2008) *Journal of American Statistical Association* **103** (482), 584-593
- Global Statistical Analysis of MISR Aerosol Data: A Massive Data Product from NASA’s TERRA Satellite
Shi, T., and Cressie, N. (2007) *Environmetrics* **18**, 665-680
- Detection of Daytime Arctic Clouds using MISR and MODIS Data
Shi, T., Clothiaux, E. E., Yu, B., Braverman, A. J. and Groff, D. N. (2007) *Remote Sensing of Environment* **107**, 172-184
- Binning in Gaussian Kernel Regularization
Shi, T. and Yu, B. (2006) *Statistica Sinica* **16**, 541-567

Conference Papers

- Data Spectroscopy: Learning Mixture Models using Eigenspaces of Convolution Operators
Shi, T., Belkin, M., and Yu, B. (2008) *Proceedings of the 25th International Conference on Machine Learning*, 936-943

- Data Mining of MISR Aerosol Product using Spatial Statistics
Shi, T., and Cressie, N. (2007) *Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining*, 712-719, IEEE press
- Fusing MISR and MODIS Information for Polar Cloud Detection
Shi, T., Yu, B., Clothiaux, E. E., and Braverman, A. J. (2004) *Proceedings of the 38th Asilomar Conference on Signals, Systems, and Computers*, vol 2, 1705-1709, IEEE press.

Preprints and Technical Reports

- Using Temporal Variability to Improve Spatial Mapping with Application to Satellite Data
Kang, E. L., Cressie, N., and **Shi, T.** (2009) *Preprint*, Department of Statistics, The Ohio State University, *submitted*
- Fixed Rank Filtering for Spatio-Temporal Data
Cressie, N., **Shi, T.**, and Kang, E. L. (2009) *Preprint*, Department of Statistics, The Ohio State University, *submitted*
- Periodontal Phenotypes in Young Adult
Chang, H.L., **Shi, T.**, Kumar, P.S. and Tatakis, D.N. (2008) *submitted*
- Data Spectroscopy: Learning Gaussian Mixture Models using Eigenspaces of Convolution Operators
Shi, T., Belkin, M., and Yu, B. (2007) *Preprint 794*, Department of Statistics, The Ohio State University.
- Cloud Detection over Snow and Ice using MISR Data
Shi, T., Yu, B., Clothiaux, E. E., and Braverman, A. J. (2004) *Technical Report 663*, Department of Statistics, University of California, Berkeley
- MISR Cloud/Ice Classification Using Linear Correlation Matching
Shi, T., Yu, B., and Braverman, A. J. (2002) *Technical Report 630*, Department of Statistics, U.C. Berkeley

Dissertation

- Polar Cloud Detection using Satellite Data with Analysis and Application of Kernel Learning Algorithms
Shi, T. (2005) *Ph.D. thesis*, University of California, Berkeley

Talks and Posters

Invited Conference Talks

- The ICSA 2009 Applied Statistics Symposium, San Francisco, CA, June 21, 2009
- Summer School of Theory and Practice of Computational Learning and Machine Learning, Chicago, IL, June 8, 2009
- The 7th World Congress in Probability and Statistics, Singapore, July 16, 2008
- MISR science team meeting, Pasadena, CA, December 7, 2007
- Joint Statistics Meeting 2007, Salt Lake City, UT, July 31, 2007
- SAMSI workshop on Geometry, Random Matrices, and Statistical Inference, Research Triangle Park, NC, January 19, 2007
- AMS/IMS/SIAM joint summer research conference: Machine Learning, Statistics, and Discovery, Snowbird, UT, June 25, 2006
- The 2nd NASA Data Mining workshop, Pasadena, CA, May 24, 2006
- NASAs TERRA Data Fusion workshop, Williamsburg, VA, August 5, 2005
- IMS/WNAR conference, Fairbanks, AL, June 22, 2005
- MISR science team meeting, Pasadena, CA, December 12, 2004
- MISR cloud workshop, Urbana-Champaign, IL, August 7, 2004
- MISR science team meeting, Pasadena, CA, December 11, 2003
- American Geophysical Union annual meeting, San Francisco, CA, December 8, 2002

Invited Seminars

- Department of Computer Science and Engineering, Ohio State University, Columbus, OH, March 6, 2009
- Department of Statistics, Rutgers University, Piscataway, NJ, October 22, 2008
- Department of Biostatistics, Yale University, New Haven, CT, September 23, 2008
- School of Statistics, University of Minnesota, Minneapolis, MN, September 18, 2008
- School of Mathematical Sciences, Peking University, Beijing, China, June 12, 2008
- Department of Statistics, Fudan University, Shanghai, China, June 3, 2008
- Department of Statistics, University of Georgia, Athens, Georgia, March 20, 2008
- National Center for Atmospheric Research, Boulder, CO, May 3, 2007
- Department of Statistics, University of California, Riverside, October, 17, 2006
- Department of Statistics, University of California, Irvine, October 16, 2006
- Summer program in Ecology and Evolution, Mathematical Bioscience Institute, the Ohio State University, Columbus OH, July 19, 2006
- Byrd Polar Research Center, the Ohio State University, Columbus, OH, January 27, 2006
- National Center for Atmospheric Research, Boulder, CO, July 12, 2005
- Department of Statistics, Pennsylvania State University, State College, PA, March 3, 2005
- Department of Statistics, The Ohio State University, Columbus, OH, February 24, 2005
- Department of Statistics, Waterloo University, Waterloo, Canada, February 17, 2005
- Bell Labs, Marry Hill, NJ, February 11, 2005
- Department of Statistics, University of Pittsburg, Pittsburg, PA, February 9, 2005
- Department of Statistics, Colorado State University, Fort Collins, CO, February 7, 2005
- Department of Statistics, Texas A&M University, College Station, TX, February 4, 2005
- Department of Statistics, University of British Columbia, Vancouver, January 31, 2005

Contributed Conference Talks

- IEEE Symposium on Computational Intelligence and Data Mining, Honolulu, HI, April 5, 2007
- MISR Data Users Science Symposium, Pasadena, CA, December 8, 2006
- Interface 2006: Massive Datasets and Streams (the 38th Symposium on the interface of statistics, computing science, and applications), Pasadena, CA, May 25, 2006

Professional Service

Service

- Associate Editor of Statistical Analysis and Data Mining (2010-2012)
- Co-instructed a short course on Data Mining and Statistical Learning at the ISCA Applied Statistics Symposium 2009
- Served on the local organization committee of the Conference on Nonparametric Statistics and Machine Learning (2010)
- Served on the program committee of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTAT 2009)
- Served on the organization committee and organized three invited paper sessions for the 2009 Applied Statistics Symposium of the International Chinese Statistical Association
- Organized and chaired an invited session on Remote Sensing in Environmental Statistics (sponsored by Section on Statistics and the Environment) for the Joint Statistical Meeting 2007
- Reviewed a grant proposal submitted to the National Science Foundation 2007

Journal Review

Biometrics [1], Computational Statistics [1], Computational Statistics and Data Analysis [2], Contemporary Mathematics [1], IEEE Transaction on Pattern Analysis and Machine Intelligence [1], Journal of Computational and Graphical Statistics [4], Journal of Machine Learning Research [8], Journal of Multivariate Analysis [1], Journal of Royal Statistics Society (Series B) [1], Machine Learning [1], Remote Sensing of Environment [3], Statistica Sinica [3]

Conference Paper Review

Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS 2009)

IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)

The SIAM International Conference on Data Mining (SCDM 2005)

The 17th Annual Conference on Learning Theory (COLT 2004)

Membership

- Member of the Institute of Mathematical Statistics
- Member of American Statistician Association
- Member of the International Chinese Statistical Association