



SLDM

## ISBIS 2014 and SLDM Joint Meeting

9-11 June 2014

Theme: Data Mining in Business and Industry

Sponsored by the  
International Society for Business and Industrial Statistics  
and the  
ASA Section on Statistical Learning and Data Mining

### Program

**Room Key:** All keynote addresses are in the Junior Ballroom. The other rooms are Meeting Rooms 1, 2, and 3, a subdivided portion of the Junior Ballroom consisting of three rooms, D1, D2, D3. The program represents these as follows:

Meeting Room 1	A
Meeting Room 2	B
Meeting Room 3	C
Division D1	D
Division D2	E
Division D3	F

## Monday, June 9

- 8:15-8:45            Registration
- 8:45-9:00           Welcoming Remarks (Junior Ballroom). David Banks, Duke University, and  
Vincenzo Esposito Vinzi, ISBIS President
- 9:00-10:00          Keynote Address  
                        Claudia Perlich, Dstillery  
                        Facing the Predictive Modeling Paradox: The Art of Making Do with  
                        the Data You Have, Not the Data You Wish You Could Get
- 10:00-10:30        Coffee Break
- 10:30-12:00        Technical Sessions
- A: Statistical Methods in Genetic Studies (org. by Hongyuan Cao, University  
of Chicago)  
                        Chair: Sounak Chakraborty, University of Missouri-Columbia  
                        Jingyun Yang, Rush University Medical Center  
                        Relation of Age to Human Brain DNA Methylation in a Cohort of  
                        Older Persons.  
                        Heping Zhang, Yale University  
                        Tree-Based Rare Variants Analyses  
                        Xiang Zhou, University of Chicago  
                        Computational Methods for Linear Mixed Effects Models in  
                        Large-Scale Genome-Wide Association Studies
- B: Record Linkage (org. and chair, Kristian Lum, Consultant)  
                        Sam Ventura, Carnegie Mellon University  
                        Large-Scale Clustering Methods with Applications in Record Linkage  
                        Beka Steorts, Carnegie Mellon University  
                        Record Linkage and Other Statistical Models for Quantifying Conflict  
                        Casualties in Syria  
                        Patrick Ball, HRDAG  
                        The Design and Implementation of a Record Linkage Pipeline
- C: Selection in Regression (org. by Yingbo Li, Clemson University)  
                        Chair: Alyson Wilson, North Carolina State University  
                        Cheryl Flynn, New York University  
                        On the Sensitivity of the Lasso to the Number of Predictor Variables  
                        Marco Ferreira, University of Missouri-Columbia  
                        Bayesian Reference Analysis for Exponential Power Regression Models  
                        Yingbo Li, Clemson University  
                        Confluent Hypergeometric Mixture of g-Priors in GLMs

## Monday, June 9

D: Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation  
(org. by Ejaz Ahmed, Brock University)

Chair: Minh Pham, SAMSI and Duke University

Xiaoli Gao, University of North Carolina–Greensboro

Complex Grouped Variable Selection

Abdulkader Hussein, University of Windsor

Estimation in High-Dimensional Spatial CAR Models

S. Ejaz Ahmed, Brock University

Big Data Analysis, Big Biases

E: Profile Monitoring (org. by Shoja'eddin Chenouri, University of Waterloo)

Chair: David Edwards, Virginia Commonwealth University

Brad Jones, SAS Institute

Augmenting a Designed Experiment with New Runs to More Precisely

Locate a Response Contour

Kamran Paynabar, Georgia Tech

A Change Point Approach for Phase-I Analysis in Multivariate

Profiles Monitoring and Diagnosis

Lee Wells, Virginia Tech

Statistical Process Control Approaches for High-Density Dimensional

Point Cloud Data-Sets

F: Business Analytics: When Academics and Practitioners Collaborate (org.  
and chair, Vincenzo Esposito Vinzi, ESSEC)

Andrew Fano, Accenture

Interactive Learning for Claims Processing

Nicolas Glady, ESSEC

A Generalized Model of Advertising: Incorporating Electronic

Word-of-Mouth into an Advertising Model

Liang Zhang, LinkedIn

Large-scale Statistical Modeling for LinkedIn Advertising

12:00-1:30

Lunch

1:30-3:00

Technical Sessions

A: Application of Advanced Algorithms (org. and chair, Fang Chen, SAS)

Jorge Silva, SAS Institute

Factorization Machines for Predictive Modeling and Recommendation

Anjishnu Banerjee, Amazon

Unstructured Algorithms to Discover Structured Patterns

Minh Pham, Statistical and Applied Mathematical Sciences Institute

Alternating Linearization for Structured Penalties

## Monday, June 9

- B: Bayesian Methods in Bioinformatics (org. by Sounak Chakraborty, University of Missouri–Columbia)  
Chair: Heping Zhang, Yale University  
Veerabhadran Baladandayuthapani, University of Texas/MD Anderson Cancer Center  
Bayesian Nonparametric Functional Models for High-Dimensional Genomics Data  
Anindya Bhadra, Purdue University  
High-Dimensional Joint Bayesian Variable and Covariance Selection: Applications in eQTL Analysis and Cancer Genomics  
Sounak Chakraborty, University of Missouri–Columbia  
Bayesian Kernel-Based Modeling and Selection of Genetic Pathways and Genes for Cancer
- C: Randomized Algorithms for Statistical Inference (org. by David Lawlor and Garvesh Raskutti, SAMSI)  
Chair: David Lawlor, SAMSI and Duke University  
Thomas Wentworth, North Carolina State University  
Leverage Score Perturbation  
Alex Gittens, eBay Research  
Dimensionality Reduction fo Large-Scale Kernel Methods  
Vikas Sindwhani, IBM T.J. Watson Research Center  
Randomization, Block Splitting and Hybrid Parallelism for Scalable Kernel Methods
- D: Analyzing Images, Speech and Clusters (org. and chair, David Banks, Duke University)  
Amita Pal, Indian Statistical Institute  
Robust Speaker Identification using Gaussian Mixture Models Based on Uncorrelated MFCC-Derived Features  
Smarajit Bose, Indian Statistical Institute  
A Hybrid Approach to Content-Based Image Retrieval  
Ilknur Kaynar Kabul, SAS Institute  
Determining the Number of Clusters in a Dataset using ABC
- E: Complex Reliability Analysis and Supersaturated Designs (org. and chair, Yili Hong, Virginia Tech)  
Kevin Wilson, University of Strathclyde  
Role of Expert Judgement in Analysing Large Complex Reliability Data Sets  
Alyson Wilson, North Carolina State University  
Combining Information to Assess System Reliability  
David Edwards, Virginia Commonwealth University  
Searching for Powerful Supersaturated Designs

Monday, June 9

F: Mining Data with Complex Structures and Constraints (org. and chair, Helen Zhang, University of Arizona)

Lingsong Zhang, Purdue University

Nested Cone Analysis Methods

Ernest Fokoué, Rochester Institute of Technology

Practical Exploration of Some Statistical and Computational Tools for Predictively Optimal Big Data Analytics

Zhaohui (Steve) Qin, Emory University

An Alternative to the Bayesian Hierarchical Model and its Application to Microarray Gene Expression Data

3:00-3:30 Coffee Break

3:30-5:00 Technical Sessions

A: Statistical Models of Risks of Chronic Diseases (org. by Marek Kimmel, Rice University)

Chair: Anindya Bhadra, Purdue University

Marek Kimmel, Rice University

Screening for Lung Cancer: Does It Work, and Who Should Be Screened?

Olga Gorlova, Dartmouth

Predicting Occult Metastasis of Lung Cancer by Modeling the Natural History and Detection

David Yankelevitz, Mount Sinai Medical Center

Lung Cancer Growth Rate Estimates in the Context of a CT Screening Program

B: Advances in Time Series Methods for Business Applications (org. and chair, Nalini Ravishanker, University of Connecticut)

Sujit Ghosh, North Carolina State University

A Computationally Efficient Flexible Observed Factor Model with Separate Dynamics for the Factor Volatilities and Correlation Matrix

Pilar Muñoz, Universitat Politècnica de Catalunya

Wind Power Forecasting and the Electricity Market

Priya Kohli, Connecticut College

Clustering Financial Time Series: A Polyspectral SLEX Approach

## Monday, June 9

- C: Emerging Trends in Bayesian Nonparametrics (org. by Anirban Bhattacharya, Texas A&M University)  
Chair: Alex Gittens, eBay Research  
Vinayak Rao, Duke University  
Dependent Probability Measures for Topic Modeling  
Guang Cheng, Purdue University  
Semiparametric Bernstein von-Mises Theorem: Second-Order Studies  
Anirban Bhattacharya, Texas A&M University  
Tensor Factorizations and Sparse Log-Linear Models
- D: Big Data Analytics in Industrial and Business Applications (org. and chair, Ming Li, SAS Institute)  
Rida Moustafa, Wal-Mart  
Big Data Analytics: An Integrated Business Solution in Wal-Mart  
Rui Zhang, IBM T. J. Watson Research Center  
A Bayesian Approach for Developing Climate Surfaces to Estimate Uncertainty in Daily Weather Interpolations  
Jin Xia, General Electric Global Research Center  
Statistical Modeling of Big, Complex Data in Industry
- E: INFORMS Session (org. and chair, Cynthia Rudin, MIT)  
Joydeep Ghosh, University of Texas–Austin  
Predictive Healthcare Analytics under Privacy Constraints  
Shawn Mankad, University of Maryland  
Analyzing Multiview Parliament Networks with Structured Matrix Factorization: Does Leadership Translate to the Twitter Universe?  
Rituparna Sen, Indian Statistical Institute, Chennai Centre  
Capital Asset Pricing Using the Horseshoe Prior
- F: Diverse Topics (org. and chair, Annie Qu, Univ. of Illinois-UC)  
George Michailidis, University of Michigan  
Estimation in High-dimensional Vector Autoregressive Models  
Junhui Wang, City University of Hong Kong  
Classification with Unstructured Predictors with an Application to Sentiment Analysis  
Xiaotong Shen, University of Minnesota  
Estimation of a Directed Acyclic Gaussian Graph
- 5:15-6:00 A: ISBIS Executive Committee and Council Meeting
- 6:00 Buses Leave for Evening Reception at SAS Institute

## Tuesday, June 10

- 9:00-10:00      Keynote Address (Junior Ballroom)  
                    Chair: Vincenzo Esposito Vinzi, ESSEC  
                    David Dunson, Duke University  
                    Big Bayes
- 10:00-10:30      Coffee Break
- 10:30-12:00      Technical Sessions
- A: ASMBI Session (org. and chair, Fabrizio Ruggeri, CNR IMATI)  
                    John Peterson, GlaxoSmithKline  
                    A Posterior Predictive Approach to Process Conformance Optimization  
                    with Complex Multivariate Models  
                    Discussant: Debdeep Pati, Florida State University  
                    Discussant: David Edwards, Virginia Commonwealth University
- B: Scan Statistics (org. by Donald Martin, North Carolina State University)  
                    Chair: Tingting Zhang, University of Virginia  
                    Nancy R. Zhang, University of Pennsylvania  
                    Scan Statistics for Detection of Genome Structural Variation  
                    Soumen N. Lahiri, North Carolina State University  
                    Resampling Approximations to the Distributions of Scan Statistics  
                    Donald E. K. Martin, North Carolina State University  
                    Distribution of Scan Statistics over Hidden States
- C: INFORMS Session (org. by Cynthia Rudin, MIT)  
                    Chair: Daniel Jeske, University of California–Riverside  
                    Edward McFowland III, Carnegie Mellon University  
                    Discovering Novel Anomalous Patterns in General Data  
                    Berk Ustun, MIT  
                    Methods and Models for Interpretable Linear Classification  
                    Paul Brooks, Virginia Commonwealth University  
                    L1-Norm Principal Component Analysis
- D: Recent Advances in Machine Learning and Data Mining (org. and chair,  
Xiaotong Shen, University of Minnesota)  
                    Yulia Baker, Rice University  
                    A General Framework for Mixed Graphical Models  
                    Adam Rothman, University of Michigan  
                    Properties of Optimizations Used in Penalized Gaussian Likelihood  
                    Inverse Covariance Matrix Estimation  
                    Yunpeng Zhao, George Mason University  
                    Maximum Likelihood Network Estimates for Social Grouping Behavior

Tuesday, June 10

- E: Topics in Industrial Statistics and Data Mining (org. and chair, Marianthi Markatou, SUNY Buffalo)  
Tom McCurdy, Stanford University and Acumen, LLC  
Updating Sequential Probability Ratio Test for Real-Time Surveillance of Vaccine Safety  
Aliza Heching, IBM T.J. Watson Research Center  
Clustering Mixed Data Subject to Measurement Error  
Ray Liu, Takeda International  
The Joint Analysis of Genomic and Pharmacological Data: A Novel Framework in Development
- F: Mining for Interactions in High-Dimensional Data (org. by Daniela Witten, University of Washington–Seattle)  
Chair: Brian Caffo, Johns Hopkins University  
Jacob Bien, Cornell University  
Mining for Interactions Using Convex Optimization  
Ali Shojaie, University of Washington  
Laplacian Shrinkage for Estimation of Inverse Covariance Matrices in Heterogenous Populations  
Han Liu, Princeton University  
Pathwise Calibrated Coordinate Descent Algorithm for Semiparametric Graph Estimation Problems: Nonconvexity with Theoretical Guarantees
- 12:00-1:30 Lunch
- 1:30-3:00 Technical Sessions
- A: Modern Decision Analysis (org. and chair, Brad Jones, SAS)  
Vincenzo Esposito Vinzi, ESSEC  
Component-Based Redundancy Path Modelling  
Xiaoming Huo, Georgia Institute of Technology  
Statistical Modelling of Bidding Prices in Online Ad Position Auctions  
Cynthia Rudin, MIT  
Interactions Between Machine Learning and Decision Making
- B: Recent Developments in High-Dimensional Data Analysis (org. and chair, Yufeng Liu, University of North Carolina–Chapel Hill)  
Yuexiao Dong, Temple University  
Model-free Variable Selection  
Chenlei Leng, University of Warwick  
High-dimensional Ordinary Least-squares Projector for Screening Variables  
Helen Zhang, University of Arizona  
Component Selection and Estimation for Functional Additive Models



Tuesday, June 10

- C: Error Control in Signal Discovery (org. by Xu Han, Temple University)  
Chair: Amita Pal, Indian Statistical Institute  
Linda Zhao, University of Pennsylvania  
    Discovering Signals with Precision Through Nonparametric Bayes  
Jichun Xie, Temple University  
    High Dimensional Tests for Brain Networks with Desirable Resolutions  
Xu Han, Temple University  
    False Discovery Control Under General Dependence
- D: Advances in Statistical Network Analysis (org. by Karl Rohe, University of Wisconsin–Madison)  
Chair: Yunpeng Zhao, George Mason University  
David Choi, Carnegie Mellon University  
    Consistency of Co-Clustering Exchangeable Graph Data  
Patrick Perry, New York University  
    Fast Hierarchical Modeling for Recommender Systems  
Norbert Binkiewicz, University of Wisconsin–Madison  
    Contextualized Spectral Network Analysis
- E: Statistical Methods for Neuroimaging Data (org. by Tingting Zhang, University of Virginia)  
Chair: Soumen Lahiri, North Carolina State University  
Brian Caffo, Johns Hopkins University  
    Analyzing Neurological Disorders Using Functional and Structural Brain Imaging Data  
Hongtu Zhu, University of North Carolina–Chapel Hill  
    Multiscale Weighted Principal Component Analysis for High-Dimensional Data on Graphs  
Tingting Zhang, University of Virginia  
    A Dynamic Directional Model for Effective Brain Connectivity Using Electroencephalographic (EEG) Time Series
- F: Classification Methods for Complex Data Structures (org. by Daniel Jeske, University of California–Riverside)  
Chair: Pilar Muñoz, Universitat Politècnica de Catalunya  
Seung Jun Shin, University of Texas/MD Anderson Cancer Center  
    Weighted Principal Support Vector Machines for Sufficient Dimension Reduction in Binary Classification  
Kush R. Varshney, IBM Thomas J. Watson Research Center  
    Talent Analytics to Predict Employee Job Roles and Skill Sets Using Diverse Data Sources  
Daniel Jeske, University of California–Riverside  
    Sequential Neutral Zone Classifiers

## Tuesday, June 10

3:00-3:30 Coffee Break

3:30-5:00 Technical Sessions

A: New Penalization Methods for Complex Data (org. by Debdeep Pati, Florida State University)

Chair: Mu Zhu, University of Waterloo

Yiyuan She, Florida State University

Group Regularized Estimation Under Strong Hierarchy

Debdeep Pati, Florida State University

Shrinkage Priors in High Dimensions

Yuan Liao, University of Maryland, College Park

Bayesian Analysis for Partially Identified Models

B: Recent Developments in Efficient Bayesian Computation (org. by Vivekananda Roy, Iowa State University)

Chair: Patrick Perry, New York University

Jorge Román, Vanderbilt University

Geometric Ergodicity of Gibbs Samplers for Bayesian Mixed Models

Aixin Tan, University of Iowa

Estimates and Standard Errors for Ratios of Normalizing Constants from Multiple Markov Chains

Vivekananda Roy, Iowa State University

Efficient Estimation and Prediction for Spatial Generalized Linear Mixed Models

C: Statistical Modeling on the Web (org. by Saharon Rosset, Tel Aviv University)

Chair: Xiaoming Huo, Georgia Institute of Technology

Ori Stitelman, Dstillery

Using Targeted Maximum Likelihood Estimation to Estimate the Impact of Online Advertising

Rick Lawrence, IBM T.J. Watson Research Center

An SMS Text Classification System for UNICEF Uganda

Melinda Han, Columbia University

Probabilistic Classification Predictions Using Aggregated Ground Truth

D: Business Analytics in Heterogenous Data (org. and chair, Hongxia Yang, IBM Watson Research Center)

Jingrui He, Steven Institute of Technology

Detecting Complex Rare Categories: Theory and Applications

Youngdeok Hwang, IBM Watson

A Statistical-Physical Approach for Air Quality Forecasting

Ban Kawas, IBM Watson

Bayesian Robust Analysis for Large Vector Autoregression Model

Tuesday, June 10

E: Estimating Conflict Mortality (org. and chair, Beka Steorts, Carnegie Mellon University)  
Shira Mitchell, Harvard University  
Population Size Estimation with Inactive Lists: Hierarchical Mixture Models and Missing Data with Application to Armed Conflict Data  
Kristian Lum, Consultant  
High-Resolution Spatio-Temporal Estimates of Undocumented Killings in a Conflict Zone  
Megan Price, HRDAG  
Record Linkage and Capture-Recapture Models for Quantifying Conflict Casualties in Syria

F: The State-of-the-Art in Classification and Subgroup Analysis (org. and chair, Xingye Qiao, SUNY Binghamton)  
Wei Sun, Purdue University  
Optimal Stability of the Nearest Neighbor Classifier  
Hanwen Huang, University of Georgia  
Multiclass Distance Weighted Discrimination  
Sijian Wang, University of Wisconsin-Madison  
Regularized Outcome Weighted Subgroup Identification for Differential Treatment Effects

5:15-5:45 A: ISBIS General Assembly Meeting

5:45-6:45 Poster Session

6:45-8:30 Banquet (Grand Ballroom II)

## Wednesday, June 11

- 9:00-10:00      Keynote Address (Junior Ballroom)  
                    Chair: Xiaotong Shen, University of Minnesota  
                    Xihong Lin, Harvard University  
                    Signal Detection in Association Analysis of Genome-wide Genetic and  
                    Genomic Data
- 10:00-10:30      Coffee Break
- 10:30-12:00      Technical Sessions
- A: INFORMS Session (org. by Cynthia Rudin, MIT)  
                    Chair: Caterina Liberati, University of Milan-Bicocca  
                    Ozden Gur Ali, Koc University  
                    Sharing Information with Behaviorally Similar Stores for Better Retail  
                    Forecasting  
                    Ram Akella, University of California–Berkeley  
                    Experimental Designs and Estimation for Online Display Advertising  
                    Attribution in Marketplaces  
                    Stephen L. France, University of Wisconsin–Mikwaukee  
                    Unsupervised Consensus Analysis for On-line Review and Questionnaire  
                    Data
- B: New Methods for Big Data (org. by Sanvesh Srivastava, SAMSI)  
                    Chair: Imad Bou-Hamad, American University of Beirut  
                    Sanvesh Srivastava, Duke University  
                    An Asynchronous Scalable Distributed Expectation-Maximization  
                    Algorithm for Massive Data: The DEM Algorithm  
                    Myung Hee Lee, Colorado State University  
                    Sparse Robust Graphical Models  
                    Taiyeong Lee, SAS Institute  
                    Incremental Response Modeling Based on Novelty Detection via  
                    One-Class Support Vector Machine
- C: Computer Experiments (org. by Abhyuday Mandal, University of Georgia)  
                    Chair: Bailey Fosdick, SAMSI and Duke University  
                    Huan Yan, Georgia Institute of Technology  
                    Calibration of Complex Computer Models: An Application in Cardiac  
                    Modeling  
                    Chunfang Devon Lin, Queen’s University  
                    Analysis of Computer Experiments with Qualitative and Quantitative  
                    Factors  
                    Lulu Kang, Illinois Institute of Technology  
                    Tapered Correlation Matrix Preconditioning for Fast Approximate  
                    Kriging

Wednesday, June 11

- D: Agent-Based Models and More (org. by Daniel Heard)  
Chair: Srinath Sampath, Hamilton Capital Management  
Daniel Heard, Duke University  
Statistical Inference Using Agent-Based Models  
Georgiy Bobashev, RTI  
myEpi: Modeling a 'Digital Self'  
Roelof Coetzer, SASOL and University of the Free State  
Classification of Dirichlet Observations applied to Industrial Problems
- E: Variable Selection (org. by Kwok Tsui, City University of Hong Kong)  
Chair: Vince Vu, Ohio State University  
John Daye, University of Arizona  
High-Dimensional Variable Screening Under Unobserved Causal Factors  
Eric Laber, North Carolina State University  
Functional Feature Construction for Personalized Treatment Regimes  
E. James Harner, West Virginia University  
Random KNN Classification with Variable Selection
- F: High Dimensions (org. by Jinchi Lv, University of Southern California)  
Chair: Andreas Artemiou, Cardiff University  
Yang Feng, Columbia University  
Model Selection in High-Dimensional Misspecified Models  
Jessie Jeng, North Carolina State University  
Weak Signal Detection in High-Dimensional Data Analysis  
Pingshou Zhong, Michigan State University  
Tests for High-Dimensional Nonparametric Functions

12:00-1:30 Lunch

1:30-3:00 Technical Sessions

- A: Theoretical and Practical Aspects of Learning in High Dimensions (org. and chair, Ernest Fokoué, Rochester Institute of Technology)  
Xingye Qiao, SUNY Binghamton  
Distance-weighted Support Vector Machine  
Mohammad Rohban, RIT and Boston University  
Efficient Distributed Topic Modeling with Provable Guarantees  
Manjari Narayan, Rice University  
Two Sample Inference on Populations of Graphical Models:  
Applications to Multi-Subject Functional Brain Connectivity

Wednesday, June 11

- B: Deep Models (org. and chair, Mu Zhu, University of Waterloo)  
Ruslan Salakhutdinov, University of Toronto  
Learning with Hierarchical Deep Models  
Hoifung Poon, Microsoft Research  
Sum-Product Networks: A New Deep Architecture  
Larry Carin, Duke University  
Deep Learning with Hierarchical Convolutional Factor Analysis
- C: Nonparametric Perspectives (org. and chair, Yichao Wu, NCSU)  
Genevera Allen, Rice University  
Sparse and Smooth Principal Components Analysis  
between Variables  
Srinath Sampath, Hamilton Capital Management  
Analysis of Agreement Between Two Long Ranked Lists  
Dungang Liu, Yale University  
Combining Nonparametric Inferences Using Data Depth and Confidence  
Distribution
- D: Network Modeling (org. and chair, Ali Shojaie, University of Washington)  
Tyler McCormick, University of Washington  
Inferring Latent Structure in Unsolicited Network Data  
Mladen Kolar, University of Chicago  
Estimating Undirected Graphs Under Weak Assumptions  
Bailey Fosdick, SAMSI and Duke University  
Relaxing Conditional Independence Assumptions in Data Fusion
- E: Data Mining in Web Technology (org. and chair, Andrew Cron,  
Weinraub Analytics)  
Brian Dalessandro, Dstillery  
On the Practical Use of Training Surrogates for Machine Learning in  
Online Display Advertising  
James Johndrow, Duke University  
Bayesian Analysis of Weighted Networks  
Olivier Coppet, Global Data Excellence  
Data mining for Onomastic Analysis
- F: Developments in Dimension Reduction (org. and chair, Seung Jun Shin,  
M.D. Anderson)  
Andreas Artemiou, Cardiff University  
Dimension Reduction Through LqSVM  
Peng Zeng, Auburn University  
Variable Selection in Multi-Index Models  
Wenxuan Zhong, University of Georgia  
Tensor Dimension Reduction

## Wednesday, June 11

3:00-3:30 Coffee

3:30-5:00 Technical Sessions

A: Innovations in Dimension Reduction of Big-Data. (org. and chair, Genevera Allen, Rice University)

Vince Vu, Ohio State

Fantope Projection and Selection

Sungkyu Jung, University of Pittsburgh

Continuum Discriminant Directions and Their High-Dimensional Asymptotic Properties

Dong Wang, University of North Carolina–Chapel Hill

Efficient Dimension Reduction for a Group of Images

B: Machine Learning and Personalized Medicine (org. and chair, Eric Laber, North Carolina State University)

Guanhua Chen, University of North Carolina–Chapel Hill

Personalized Dose Finding Using Outcome Weighted Learning

Daniel Lizotte, University of Waterloo

Set-Valued Dynamic Treatment Regimes for Competing Outcomes

Yingqi Zhao, University of Wisconsin

Tree-Based Methods for Optimal Treatment Allocation

C: Active Learning (org. and chair, Ann Lee, Carnegie Mellon University)

Aarti Singh, Carnegie Mellon University

Active Completion of Coherent Low-Rank Matrices and Tensors

Jarvis Haupt, University of Minnesota

Adaptive Compressive Sensing of Signals Exhibiting Tree-structured Sparsity: Fundamental Limits, Implications, and Applications

Steve Hanneke, Free Spirit

Minimax Analysis of Active Learning

D: Theoretical and Computational Foundations of Big Data Analysis (org. and chair, Han Liu, Princeton)

Le Song, Georgia Institute of Technology

Nonparametric Latent Variable Models via Kernel Embedding of Distributions

Sivaranman Balakrishnan, UC Berkeley

A case study of the EM algorithm

David Banks, Duke University

Text Mining in a Blog Network

Wednesday, June 11

E: Statistical Learning with Massive and Complex Data (org. by Ji Zhu, University of Michigan)

Chair: Tyler McCormick, University of Washington

Yingying Fan, University of Southern California

Asymptotic Equivalence of Regularization Methods in Thresholded Parameter Space

Xi Luo, Brown University

Network Based Discriminant Analysis with Applications to fMRI

Hua Zhou, North Carolina State University

Online Statistical Learning Algorithms

F: Data Mining in Commerce (org. by David Banks, Duke University)

Chair: Daniel Heard, Duke University

Imad Bou-Hamad, American University of Beirut

Credit Risk Assessment: A Random Forest Based Approach

Caterina Liberati, University of Milano-Bicocca

Visualization and Measuring of Dynamic Customer Satisfaction:  
A Banking Case

Federico Armentano, Universidad Nacional de Rosario

Comparison of Models for Time Series Forecasts of Tourism in Argentina