

# Interface 2012 Program Book

43<sup>rd</sup> Symposium on the Interface of  
Computing Science and Statistics

Theme:

Future of Statistical Computing:  
Internet Scale Data, Flexible Modeling, and Visualization

May 16-18, 2012

Wednesday 8 am - Friday noon

Program Co-Chairs:

David W Scott, Rice University  
Hadley Wickham, Rice University  
Jeffrey Morris, MD Anderson

(Typeset May 18 at 1:00 pm)

# Wednesday, May 16, 2012

7:30 am - 5:00 pm                      **Registration**                      Lobby

8:00 am - 9:45 am                      **Welcome and Keynote Address**                      Auditorium

- *Introductions* David Scott, Rice University
- *Welcoming Remarks* Ned Thomas, Dean of Engineering, Rice University
- *Matrix Completion and Large-Scale SVD Computations*, Trevor Hastie, Stanford University

9:45 am - 10:15 am                      **Morning Break**                      Lobby

10:15 am - 12:00 noon                      **Technical Sessions**

• **Software Development in R**                      Auditorium

Organizer: Duncan Murdoch, University of Western Ontario

1. *RStudio - Integrated Development Environment for R*, JJ Allaire, Founder RStudio Project
2. *Efficient R Parallel Loops on Long-Latency Platforms*, Norm Matloff, University of California Davis
3. *Debugging Support in R*, Duncan Murdoch, University of Western Ontario

• **Using Statistical Modeling for Pricing Applications**                      Room 1064

Organizer: John Salch, PROS Revenue Management, Inc

1. *The Complexity of Petroleum Pricing at PROS*, Daniel Covarrubias, PROS Revenue Management, Inc
2. *Segmentation for Extremely Large Datasets*, Evan Brott, PROS Revenue Management, Inc
3. *A New Approach for Forecasting Demand via Willingness to Pay*, Ed Kambour, PROS Revenue Management, Inc

- **High-Dimensional Graphical Models** Room 1070  
 Organizer: Genevera Allen, Baylor College of Medicine and Rice University
  1. *A Log-Linear Graphical Model*, Genevera Allen, Baylor College of Medicine and Rice University
  2. *Dynamic Logistic Regression and Dynamic Model Averaging for Binary Classification*, Tyler McCormick, University of Washington
  3. *Greedy Algorithms for Learning Discrete and Gaussian Graphical Models*, Pradeep Ravikumar, University of Texas Austin

12:00 noon - 1:30 pm Lunch

1:30 pm - 3:15 pm Technical Sessions

- **Man AND Machine: the Conversation, using the Language of Interactive Graphics** Auditorium  
 Organizer: Heike Hofmann, Iowa State University
  1. *Interactive Graphics in R: The Plumbing and the Painting*, Michael Lawrence, Genentech
  2. *cranvas: Building from Plumbing and Painting*, Yihui Xie, Iowa State University
  3. *Enhancing Web Pages with R in the Browser*, Gabriel Becker, University of California-Davis

- **Unifying Statistical Sciences** Room 1064  
 Organizer: Arnie Goodman, Collaborative Data Solutions
  1. *Modeling, Dependence, Classification, United Statistical Science, Many Cultures*, Manny Parzen, Texas A&M University
  2. *Discussion*, Don Ylvisaker, UCLA
  3. *Discussion*, Joe Newton, Texas A&M University

- **Tensors: Decompositions and Applications** Room 1070  
 Organizer: Eric Chi, UCLA
  1. *Nonnegative Tensor Factorizations*, Eric Chi, University of California, Los Angeles

2. *Tensor Regression with Applications in Neuroimaging Data Analysis*, Hua Zhou, North Carolina State University
3. *GSVD Comparison of Cancer Patient-Matched Genomic Profiles Predicts Survival and Novel Drug Targets*, Preethi Sankaranarayanan\* and Orly Alter, University of Utah

• **Late Breaking Session** Room 1075

Organizer: David Scott, Rice University

1. *From Single-SNP to Wide-Locus: Increasing Resolution and Power of GWAS*, Knut M. Wittkowski\*, Rockefeller University; Vikas Sonakya, Rockefeller University; Tingting Song, Rockefeller University; Martin P. Seybold, Stuttgart University; Mehdi Keddache, Cincinnati Children's Hospital Medical Center; and Kartina Durner, Mount Sinai School of Medicine
2. *Multivariate Spatial Process Modeling: An Overview*, William Kleiber, UCAR
3. *Uncertainty in Regional Climate Experiments*, Steve Sain, UCAR

3:15pm - 3:45 pm Afternoon Break Lobby

3:45 pm - 5:30 pm Technical Sessions

• **Statistical Models for Complex Functional Data** Auditorium

Organizer: Veera Baladandayuthapani, UT MD Anderson Cancer Center

1. *Regression Modeling with Images as Predictors*, Todd Ogden, Columbia University
2. *Efficient Spatial Smoothing Over Irregular Domains Using Functional PCA*, Lan Zhou, Texas A&M University
3. *Bayesian Nonparametric Functional Models for High-Dimensional Genomics Data*, Veera Baladandayuthapani, UT MD Anderson Cancer Center

• **Some Data Based Analyses in Real World Finance** Room 1064

Organizer: James Thompson, Rice University

1. *Estimating the Term Structure With a Semi-Parametric Bayesian Population Model: An Application to Corporate Bonds and Ratings*, Katherine B. Ensor\*, Rice University; Alejandro Cruz-Marcello, Capital One Bank; and Gary Rosner, Johns Hopkins University

2. *Model Specification Error with High-Speed Computational Propagation as Seen in the Subprime Meltdown*, John A. Dobelman, Rice University
3. *Empirical Data Based Alternatives to Classical Techniques in Portfolio Formation*, James R. Thompson, Rice University

• **Bayesian Multiple Comparison Procedures** Room 1070

Organizer: Peter Mueller, University of Texas at Austin

1. *A Bayesian Discovery Procedure*, Michele Guindani, M.D. Anderson Cancer Center
2. *Bayesian Multiplicity Control for RNA-Seq Data on Differential Expression Using Gene Ontology Information*, David B. Dahl, Texas A&M University
3. *Bayesian Decision Theoretic Multiple Comparison Procedures: An Application to Phage Display Data*, Luis Leon, University of Florida

5:30pm - 7:30 pm    **Conference Reception and Happy Hour**    Duncan Hall Lobby

6:00pm - 7:30 pm    **Board Meeting (By Invitation)**    1049 Duncan Hall

## Thursday, May 17, 2012

8:15 am - 10:00 am    **Technical Sessions**

• **JCGS Highlights at the Interface** Auditorium

Organizer: Richard Levine, JCGS Editor, San Diego State University

1. *Symbolic-Covariance Principal Component Analysis and Visualization for Interval-Valued Data*, Jennifer Le-Rademacher, Medical College of Wisconsin; and Lynne Billard, University of Georgia
2. *Local Derivative-Free Approximation of Computationally Expensive Posterior Densities*, Nikolay Bliznyuk, University of Florida
3. *Fitting Social Network Models Using Varying Truncation Stochastic Approximation MCMC Algorithm*, Faming Liang, Texas A&M University

• **Automatic, Flexible Computational Methods with Applications in Biostatistics** Room 1064

Organizer: David Scott, Rice University

1. *Statistical Analysis of Computer Algorithms for Assigning Cause-of-Death Codes*, Diba Khan\*, Centers for Disease Control & Prevention; National Center for Health Statistics; Myron Katzoff, Centers for Disease Control & Prevention; National Center for Health Statistics; Charles Sirc, Centers for Disease Control & Prevention; National Center for Health Statistics; Donna L. Hoyert, Centers for Disease Control & Prevention; National Center for Health Statistics; Alaina Elliott, Centers for Disease Control & Prevention; National Center for Health Statistics
2. *Bayesian Survival Trees for Clustered Observations with Application to Tooth Prognosis*, Richard Levine, San Diego State University
3. *Looking Beyond the Lamppost: Flexible Methods Bringing Light into the Dark Alleys of Complex Data*, Jeffrey S. Morris, The University of Texas M.D. Anderson Cancer Center

• **Information Mining** Room 1070

Organizer: William Szewczyk, National Security Agency

1. *Choosing a Dissimilarity for Classification*, Adam Cardinal-Stakenas, Department of Defense
2. *Strategies for Streaming Exploratory Data Analysis*, William Szewczyk, National Security Agency
3. *Solving a Story with Multiple Unknowns*, Andy Frenkiel, IBM/T.J. Watson Research

• **Contributed Paper Session I** Room 1075

Organizer: David Scott, Rice University

1. *Optimal Reduced Isotonic Regression*, Janis Hardwick and Quentin Stout\*, University of Michigan
2. *Exponential-Family Random Network Models*, Ian Fellows\* and Mark S. Handcock, University of California, Los Angeles
3. *A Projection Pursuit Index Based on Kernel PCA with Gaussian Kernels*, Victor Muniz\*, Johan Van Horebeek, and Rogelio Ramos, Research Center in Mathematics. Monterrey, Mexico

4. *Dependent Pólya Urn Schemes*, Bernardo Nipoti, University of Texas MD Anderson Cancer Center

10:00 am - 10:30 am                      **Morning Break**                      Lobby

10:30 am - 12:15 pm                      **Technical Sessions**

• **Big, Fast, and Interactive Data**                      Auditorium

Organizer: Michael Kane, Yale University

1. *Elastic Computing with R and Redis*, Bryan Lewis, Paradigm4
2. *How Google Estimates Traffic for Millions of Queries*, Tim Hesterberg, Google
3. *EDA, Visualization and Collaboration on the Web*, Carlos Scheidegger, AT&T Labs

• **Bayesian Hierarchical Models for High Dimensional Spatial Data**      Room 1064

Organizer: Sudipto Banerjee, University of Minnesota

1. *Space-Time Data Fusion Under Error in Computer Model Output: An Application to Modeling Air Quality*, Veronica J. Berrocal, University of Michigan
2. *An Adaptive Spatial Model for Precipitation Data From Multiple Satellites Over Large Regions*, Avishek Chakraborty, Texas A&M University
3. *Flexible Predictive Process Spatial Factor Models for Misaligned Data Sets*, Qian Ren and Sudipto Banerjee, University of Minnesota

• **Computing in Statistics Education**                      Room 1070

Organizer: Webster West, Texas A&M University

1. *Teaching Formulas in Statistics Classes: When Is It Beneficial?*, David Lane, Rice University
2. *Using Simulations to Teach Statistical Inference*, Beth Chance, California Polytechnic
3. *The Impact of Technology on the Teaching of Statistics*, Webster West, Texas A&M University

- **Contributed Paper Session II** Room 1075  
 Organizer: David Scott, Rice University
  1. *A Split-and-Conquer Approach for Analysis of Extraordinarily Large Data*, Xueying Chen\* and Minge Xie, Rutgers University
  2. *Manipulating Dates and Times in R With the Lubridate Package*, Garrett Grolemund\* and Hadley Wickham, Rice University
  3. **mpoly**: *Multivariate Polynomials in R*, David Kahle, Baylor University

12:15 pm - 2:00 pm Lunch

2:00 pm - 3:45 pm Technical Sessions

- **Inference on Graphs** Auditorium  
 Organizer: Organizers: Carey Priebe and Dave Marchette, Johns Hopkins University and Naval Surface Warfare Center
  1. *Graph Inference with Imperfect Edge Classifiers*, Michael Trosset, Indiana University
  2. *Consistent Embedding of Stochastic Blockmodels*, Minh Tang, Johns Hopkins University
  3. *Title Vertex Nomination: Improved Fusion of Content and Context*, Glen Copersmith, Johns Hopkins University

- **Statistical and Computational Methods for Large Spatial Data Sets** Room 1064

Organizer: Jianhua Huang, Texas A&M University

1. *Covariance Decomposition with Low Rank and Sparse Representation for Large Spatial Datasets*, Huiyan Sang, Texas A&M University
2. *Nonstationary Cross-Covariance Models for Multivariate Processes on a Globe*, Mikyoung Jun, Texas A&M University
3. *Bayesian Estimation for Large Spatial Datasets Observed on a Lattice*, Jonathan R. Stroud, George Washington University

- **Woman VS Machine: The Inference Battle** Room 1070  
 Organizer: Di Cook, Iowa State University



1. *Statistical Inference after Model Selection*, Andreas Buja, Wharton School, University of Pennsylvania
2. *Facing Off: Power of Visual and Classical Tests*, Heike Hofmann, Iowa State University
3. *Turk Experiments for Visual Inference*, Mahbub Majumder, Iowa State University

• **Contributed Paper Session III**

Room 1075

Organizer: David Scott, Rice University

1. *Bayesian Multiplicity Control for Graphs*, Riten Mitra\*, University of Texas MD Anderson Cancer Center; Peter Mueller, University of Texas at Austin; and Yuan Ji, University of Texas MD Anderson Cancer Center
2. *A Nonparametric Bayesian Model for Local Clustering*, Juhee Lee, University of Texas MD Anderson Cancer Center
3. *Testing Goodness of Fit of Protein Conformational Sampling Models*, Mehdi Maadooliat, IAMCS, Texas A&M University
4. *Improving Compton Scatter Camera Image Resolution Using Classification and the Metropolis-Hastings Algorithm*, Dennis Mackin, University of Texas MD Anderson Cancer Center

3:45pm - 4:15 pm                      **Afternoon Break**                      Lobby

4:15 pm - 5:30 pm                      **Banquet Keynote**                      Auditorium

- *Repetition and Surprise, Rehearsal and Reinvention*, Mark Hansen, UCLA

5:30 pm - 6:30 pm                      **Reception**                      Lobby

6:30 pm - 9:00 pm                      **Banquet**                      Lobby

# Friday, May 18, 2012

8:15 am - 10:00 am

## Technical Sessions

- **Modeling, Analyzing, and Visualizing High-Dimensional Data in Genomics**  
Auditorium

Organizer: Karen Kafadar, Indiana University

1. *Fast Graphical Model Estimation and Its Applications*, Daniela Witten, University of Washington
2. *Conditional Network Testing in High-dimensional Dependent Data*, Gary Gadbury, Kansas State University
3. *Robust Identification of Conditional Gene Expression in Development of Onthophagus Beetles*, Guilherme V Rocha, Indiana University

- **Visualization and Computational Methods for Actigraphy Data** Room 1064

Organizer: Jürgen Symanzik, Utah State University

1. *Powerful Actigraphy Data Through Functional Representation*, Jimin Ding, Washington University of St. Louis
2. *Reliability and Reproducibility Issues in Accelerometer-Based Estimates of Physical Activity*, Julia Kozlitina, UT Southwestern Medical Center
3. *Movelets: A Dictionary of Movement*, Jeff Goldsmith, Johns Hopkins Bloomberg School of Public Health

- **Developing Intelligence in Unmanned Ground Systems** Room 1070

Organizer: Barry Bodt, U.S. Army Research Laboratory

1. *Some Thoughts on Experimentation Philosophy in the Robotics CTA*, Barry Bodt, U.S. Army Research Laboratory
2. *Preliminary Performance Evaluation of Autonomous Mobility in Small UGVs*, Alberto Lacaze, Robotic Research, LLC
3. *Using Expectations to Drive Cognitive Behavior*, Unmesh Kurup, Carnegie Mellon University

10:00 am - 10:15 am Morning Break Lobby

10:15 am - 12:00 noon Technical Sessions

- **Computational Tools and Statistical Methods with Medical Applications**  
Auditorium

Organizer: Bradley Broom and Kim-Anh Do, University of Texas M.D. Anderson Cancer Center

1. *Graph-Based Signal Integration for High-Throughput Phenotyping*, Jorge Hershkovic\*, University of Texas M.D. Anderson Cancer Center and Elmer Bernstam, UT Health Sciences Center at Houston
2. *Extending the Grammar of Graphics for Genomic Data: an R Implementation*, Tengfei Yin, Iowa State University; Dianne Cook\*, Iowa State University and Michael Lawrence, Genentech
3. *Massive Parallelization of Serial Inference Algorithms for a Complex Generalized Linear Model*, Marc Suchard, UCLA; Shawn Simpson, Columbia University; Ivan Zorych, Columbia University; Patrick Ryan, Johnson & Johnson Pharmaceutical Research and Development; and David Madigan\*, Columbia University

- **Applications of Interactive Graphics in R** Room 1064

Organizer: David Scott, Rice University

1. *Using R and Web Technologies to Create Analytics and Apps: Part of Making Statistics Relevant in a Large Organization*, Chad Shaw, Baylor College of Medicine
2. *Exploring Statistical Strategies for Use in Challenge-Response Experiments*, Matthew S. Shotwell\*, Kenneth J. Drake, Veniamin Y. Sidorov, and John P. Wikswo: Vanderbilt University School of Medicine
3. *The Anscombe Data Sets: Explained and Expanded*, Jürgen Symanzik, Utah State University

- **Generalized Parallel Coordinates** Room 1070

Organizer: Rida Moustafa, George Washington University & dMining Technology

1. *Cluster Detection and Visualization with Generalized Parallel Coordinates*, Rida Moustafa, George Washington University & dMining Technology

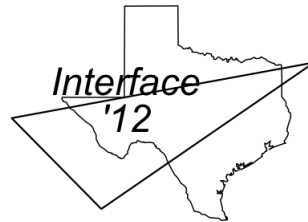
2. *Visual Cluster and Outlier Detection with L-plot*, Michael D. Larsen\*, George Washington University & dMining Technology; Rida E. Moustafa, George Washington University & dMining Technology; and Ali S. Hadi, American University in Cairo & dMining Technology
3. *Visual Analytics Approach for Social Network Interaction*, Jie Cong, George Washington University & dMining Technology and Rida Moustafa, George Washington University & dMining Technology

• **Contributed Paper Session IV**

Room 1075

Organizer: David Scott, Rice University

1. *Parallel Monte Carlo Simulation for the Sensitivity Analysis of Expected Shortfall by Means of a Second-Order Approximation*, Güven Gül Polat, Istanbul Technical University
2. *Comparison of Binary Discrimination Methods for High Dimension Low Sample Size Data*, Addy Bolivar-Cime, Rice University and J.S. Marron, University of North Carolina, Chapel Hill
3. *Factor Model for Forecasting with Multi-collinearity and Nonlinear Dependence*, Joseph Egbulefu, Rice University
4. *Relations Between Attentional Structure and Attentional Function: Utilization of Alternative Statistical Approaches*, Paulina Kulesz, University of Houston



## Index

- Allaire, JJ, 2  
Allen, Genevera, 3  
Alter, Orly, 4
- Baladandayuthapani, Veera, 4  
Banerjee, Sudipto, 7  
Becker, Gabriel, 3  
Bernstam, Elmer, 11  
Berrocal, Veronica, 7  
Billard, Lynne, 5  
Bliznyuk, Nikolay, 5  
Bodt, Barry, 10  
Bolivar-Cime, Addy, 12  
Broom, Bradley, 11  
Brott, Evan, 2  
Buja, Andreas, 9
- Cardinal-Stakenas, Adam, 6  
Chakraborty, Avishek, 7  
Chance, Beth, 7  
Chen, Xueying, 8  
Chi, Eric, 3  
Cong, Jie, 12  
Cook, Di, 8, 11  
Coppersmith, Glen, 8  
Covarrubias, Daniel, 2  
Cruz-Marcello, Alejandro, 4
- Dahl, David, 5  
Ding, Jimin, 10  
Do, Kim-Anh, 11  
Dobelman, John, 5  
Drake, Kenneth, 11  
Durner, Martina, 4
- Egbulefu, Joseph, 12  
Elliott, Alaina, 6  
Ensor, Katherine, 4
- Fellows, Ian, 6
- Frenkiel, Andy, 6
- Gadbury, Gary, 10  
Goldsmith, Jeff, 10  
Goodman, Arnie, 3  
Grolemund, Garrett, 8  
Guindani, Michele, 5
- Hadi, Ali, 12  
Handcock, Mark, 6  
Hansen, Mark, 9  
Hastie, Trevor, 2  
Herskovic, Jorge, 11  
Hesterberg, Tim, 7  
Hofmann, Heike, 3, 9  
Hoyert, Donna, 6  
Huang, Jiahua, 8
- Ji, Yuan, 9  
Jun, Mikyoung, 8
- Kafadar, Karen, 10  
Kahle, David, 8  
Kambour, Ed, 2  
Kane, Michael, 7  
Katzoff, Myron, 6  
Keddache, Mehdi, 4  
Khan, Diba, 6  
Kleiber, William, 4  
Kozlitina, Julia, 10  
Kulesz, Paulina, 12  
Kurup, Unmesh, 10
- Lacaze, Alberto, 10  
Lane, David, 7  
Larsen, Michael, 12  
Lawrence, Michael, 3, 11  
Le-Rademacher, Jennifer, 5  
Lee, Juhee, 9  
Leon, Luis, 5

Levine, Richard, 5, 6  
 Lewis, Bryan, 7  
 Liang, Faming, 5  
  
 Maadooliat, Mehdi, 9  
 Mackin, Dennis, 9  
 Madigan, David, 11  
 Majumder, Mahbub, 9  
 Marchette, Dave, 8  
 Marron, Steve, 12  
 Matloff, Norm, 2  
 McCormick, Tyler, 3  
 Mitra, Riten, 9  
 Morris, Jeff, 1, 6  
 Moustafa, Rida, 11, 12  
 Mueller, Peter, 5, 9  
 Muniz, Victor, 6  
 Murdoch, Dunca, 2  
 Murdoch, Duncan, 2  
  
 Newton, Joe, 3  
 Nipoti, Bernardo, 7  
  
 Ogden, Todd, 4  
  
 Parzen, Manny, 3  
 Polat, Güven, 12  
 Priebe, Carey, 8  
  
 Ramos, Rogelio, 6  
 Ravikumar, Pradeep, 3  
 Ren, Qian, 7  
 Rocha, Guilherme, 10  
 Rosner, Gary, 4  
 Ryan, Patrick, 11  
  
 Sain, Steve, 4  
 Salch, John, 2  
 Sang, Huiyan, 8  
 Sankaranarayanan, Preethi, 4  
 Scheidegger, Carlos, 7  
 Scott, David, 1, 2, 4, 6, 8, 9, 11, 12  
 Seybold, Martin, 4  
  
 Shaw, Chad, 11  
 Shotwell, Matthew, 11  
 Sidorov, Veniamin, 11  
 Simpson, Shawn, 11  
 Sirc, Charles, 6  
 Sonakya, Vikas, 4  
 Song, Tingting, 4  
 Stout, Quentin, 6  
 Stroud, Jonathan, 8  
 Suchard, Marc, 11  
 Symanzik, Jürgen, 10, 11  
 Szewczyk, William, 6  
  
 Tang, Minh, 8  
 Thomas, Ned, 2  
 Thompson, James, 4, 5  
 Trosset, Michael, 8  
  
 Van Horebeek, Johan, 6  
  
 West, Webster, 7  
 Wickham, Hadley, 1, 8  
 Wikswo, John, 11  
 Witten, Daniela, 10  
 Wittkowski, Knut, 4  
  
 Xie, Minge, 8  
 Xie, Yihui, 3  
  
 Yin, Tengfei, 11  
 Ylvisaker, Don, 3  
  
 Zhou, Hua, 4  
 Zhou, Lan, 4  
 Zorych, Ivan, 11