

# **43rd Symposium on the Interface of Computing Science and Statistics 2012 (Interface 2012)**

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

**Abstracts**

**Houston, Texas, USA  
16-18 May 2012**

**ISBN: 978-1-62276-719-9**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2012) by the Interface Foundation of North America  
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Interface Foundation of North America  
at the address below.

Interface Foundation of North America  
PO Box 7460  
Fairfax Station, VA 22039

Phone: (703) 993-1212  
Fax: (703) 993-1700

[interface@galaxy.gmu.edu](mailto:interface@galaxy.gmu.edu)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

## KEYNOTE ADDRESS

<b>Matrix Completion and Large-Scale SVD Computations</b> .....	1
<i>Trevor Hastie</i>	

## SOFTWARE DEVELOPMENT IN R

<b>RStudio - Integrated Development Environment for R</b> .....	1
<i>JJ Allaire</i>	
<b>Efficient R Parallel Loops on Long-Latency Platforms</b> .....	2
<i>Norm Matloff</i>	
<b>Debugging Support in R</b> .....	2
<i>Duncan Murdoch</i>	

## USING STATISTICAL MODELING FOR PRICING APPLICATIONS

<b>The Complexity of Petroleum Pricing at PROS</b> .....	2
<i>Daniel Covarrubias</i>	
<b>Segmentation for Extremely Large Datasets</b> .....	2
<i>Evan Brott</i>	
<b>A New Approach for Forecasting Demand via Willingness to Pay</b> .....	3
<i>Ed Kambour</i>	

## HIGH-DIMENSIONAL GRAPHICAL MODELS

<b>A Log-Linear Graphical Model</b> .....	3
<i>Genevera Allen</i>	
<b>Dynamic Logistic Regression and Dynamic Model Averaging for Binary Classification</b> .....	3
<i>Tyler McCormick</i>	
<b>Greedy Algorithms for Learning Discrete and Gaussian Graphical Models</b> .....	4
<i>Pradeep Ravikumar</i>	

## MAN AND MACHINE: THE CONVERSATION, USING THE LANGUAGE OF INTERACTIVE GRAPHICS

<b>Interactive Graphics in R: The Plumbing and the Painting</b> .....	5
<i>Michael Lawrence</i>	
<b>cranvas: Building from Plumbing and Painting</b> .....	5
<i>Yihui Xie</i>	
<b>Enhancing Web Pages with R in the Browser</b> .....	6
<i>Gabriel Becker</i>	

## UNIFYING STATISTICAL SCIENCES

<b>Modeling, Dependence, Classification, United Statistical Science, Many Cultures</b> .....	6
<i>Manny Parzen</i>	

## TENSORS: DECOMPOSITIONS AND APPLICATIONS

<b>Nonnegative Tensor Factorizations</b> .....	7
<i>Eric Chi</i>	

<b>Tensor Regression with Applications in Neuroimaging Data Analysis</b> .....	7
<i>Hua Zhou</i>	
<b>GSVD Comparison of Cancer Patient-Matched Genomic Profiles Predicts Survival and Novel Drug Targets</b> .....	8
<i>Preethi Sankaranarayanan, Orly Alter</i>	

**LATE BREAKING SESSION**

<b>From Single-SNP to Wide-Locus: Increasing Resolution and Power of GWAS</b> .....	8
<i>Knut M. Wittkowski, Vikas Sonakya, Tingting Song, Martin P. Seybold</i>	
<b>Multivariate Spatial Process Modeling: An Overview</b> .....	9
<i>William Kleiber</i>	
<b>Uncertainty in Regional Climate Experiments</b> .....	9
<i>Steve Sain</i>	

**STATISTICAL MODELS FOR COMPLEX FUNCTIONAL DATA**

<b>Regression Modeling with Images as Predictors</b> .....	10
<i>Todd Ogden</i>	
<b>Efficient Spatial Smoothing Over Irregular Domains Using Functional PCA</b> .....	10
<i>Lan Zhou</i>	
<b>Bayesian Nonparametric Functional Models for High-Dimensional Genomics Data</b> .....	10
<i>Veera Baladandayuthapani</i>	

**SOME DATA BASED ANALYSES IN REAL WORLD FINANCE**

<b>Estimating the Term Structure With a Semi-Parametric Bayesian Population Model: An Application to Corporate Bonds and Ratings</b> .....	11
<i>Katherine B. Ensor, Alejandro Cruz-Marcello, Gary Rosner</i>	
<b>Model Specification Error with High-Speed Computational Propagation as Seen in the Subprime Meltdown</b> .....	11
<i>John A. Dohelman</i>	
<b>Empirical Data Based Alternatives to Classical Techniques in Portfolio Formation</b> .....	12
<i>James R. Thompson</i>	

**BAYESIAN MULTIPLE COMPARISON PROCEDURES**

<b>A Bayesian Discovery Procedure</b> .....	12
<i>Michele Guindani</i>	
<b>Bayesian Multiplicity Control for RNA-Seq Data on Differential Expression Using Gene Ontology Information</b> .....	12
<i>David B. Dahl</i>	
<b>Bayesian Decision Theoretic Multiple Comparison Procedures: An Application to Phage Display Data</b> .....	13
<i>Luis Leon</i>	

**JCGS HIGHLIGHTS AT THE INTERFACE**

<b>Symbolic-Covariance Principal Component Analysis and Visualization for Interval-Valued Data</b> .....	14
<i>Jennifer Le-Rademacher, Lynne Billard</i>	
<b>Local Derivative-Free Approximation of Computationally Expensive Posterior Densities</b> .....	14
<i>Nikolay Bliznyuk</i>	
<b>Fitting Social Network Models Using Varying Truncation Stochastic Approximation MCMC Algorithm</b> .....	14
<i>Faming Liang</i>	

## **AUTOMATIC, FLEXIBLE COMPUTATIONAL METHODS WITH APPLICATIONS IN BIO-STATISTICS**

<b>Statistical Analysis of Computer Algorithms for Assigning Cause-of-Death Codes</b> .....	15
<i>Diba Khan, Myron Katzo, Charles Sirc, Donna L. Hoyert, Alaina Elliott</i>	
<b>Bayesian Survival Trees for Clustered Observations with Application to Tooth Prognosis</b> .....	16
<i>Richard Levine</i>	
<b>Looking Beyond the Lamppost: Flexible Methods Bringing Light into the Dark Alleys of Complex Data</b> .....	16
<i>Jeffrey S. Morris</i>	

## **INFORMATION MINING**

<b>Choosing a Dissimilarity for Classification</b> .....	17
<i>Adam Cardinal-Stakenas</i>	
<b>Strategies for Streaming Exploratory Data Analysis</b> .....	17
<i>William Szewczyk</i>	
<b>Solving a Story with Multiple Unknowns</b> .....	18
<i>Andy Frenkiel</i>	

## **CONTRIBUTED PAPER SESSION I**

<b>Optimal Reduced Isotonic Regression</b> .....	18
<i>Janis Hardwick, Quentin Stout</i>	
<b>Exponential-Family Random Network Models</b> .....	19
<i>Ian Fellows, Mark S. Handcock</i>	
<b>A Projection Pursuit Index Based on Kernel PCA with Gaussian Kernels</b> .....	19
<i>Victor Muniz, Johan Van Horebeek, Rogelio Ramos</i>	
<b>Dependent Polya Urn Schemes</b> .....	19
<i>Bernardo Nipoti</i>	

## **BIG, FAST, AND INTERACTIVE DATA**

<b>Elastic Computing with R and Redis</b> .....	20
<i>Bryan Lewis</i>	
<b>How Google Estimates Traffic for Millions of Queries</b> .....	20
<i>Tim Hesterberg</i>	
<b>EDA, Visualization and Collaboration on the Web</b> .....	20
<i>Carlos Scheidegger</i>	

## **BAYESIAN HIERARCHICAL MODELS FOR HIGH DIMENSIONAL SPATIAL DATA**

<b>Space-Time Data Fusion Under Error in Computer Model Output: An Application to Modeling Air Quality</b> .....	21
<i>Veronica J. Berrocal</i>	
<b>An Adaptive Spatial Model for Precipitation Data From Multiple Satellites Over Large Regions</b> .....	21
<i>Avishek Chakraborty</i>	
<b>Flexible Predictive Process Spatial Factor Models for Misaligned Data Sets</b> .....	22
<i>Qian Ren, Sudipto Banerjee</i>	

## **COMPUTING IN STATISTICS EDUCATION**

<b>Teaching Formulas in Statistics Classes: When Is It Beneficial?</b> .....	23
<i>David Lane</i>	
<b>Using Simulations to Teach Statistical Inference</b> .....	23
<i>Beth Chance</i>	

<b>The Impact of Technology on the Teaching of Statistics</b> .....	23
<i>Webster West</i>	

**CONTRIBUTED PAPER SESSION II**

<b>A Split-and-Conquer Approach for Analysis of Extraordinarily Large Data</b> .....	23
<i>Xueying Chen, Minge Xie</i>	
<b>Manipulating Dates and Times in R With the Lubridate Package</b> .....	24
<i>Garrett Grolemund, Hadley Wickham</i>	
<b>mpoly: Multivariate Polynomials in R</b> .....	24
<i>David Kahle</i>	

**INFERENCE ON GRAPHS**

<b>Graph Inference with Imperfect Edge Classifiers</b> .....	25
<i>Michael Trosset</i>	
<b>Consistent Embedding of Stochastic Blockmodels</b> .....	25
<i>Minh Tang</i>	
<b>Title Vertex Nomination: Improved Fusion of Content and Context</b> .....	25
<i>Glen Coppersmith</i>	

**STATISTICAL AND COMPUTATIONAL METHODS FOR LARGE SPATIAL DATA SETS**

<b>Covariance Decomposition with Low Rank and Sparse Representation for Large Spatial Datasets</b> .....	26
<i>Huiyan Sang</i>	
<b>Nonstationary Cross-Covariance Models for Multivariate Processes on a Globe</b> .....	26
<i>Mikyounng Jun</i>	
<b>Bayesian Estimation for Large Spatial Datasets Observed on a Lattice</b> .....	27
<i>Jonathan R. Stroud</i>	

**WOMAN VS MACHINE: THE INFERENCE BATTLE**

<b>Statistical Inference after Model Selection</b> .....	27
<i>Andreas Buja</i>	
<b>Facing Off: Power of Visual and Classical Tests</b> .....	28
<i>Heike Hofmann</i>	
<b>Turk Experiments for Visual Inference</b> .....	28
<i>Mahbub Majumder</i>	

**CONTRIBUTED PAPER SESSION III**

<b>Bayesian Multiplicity Control for Graphs</b> .....	28
<i>Riten Mitra, Peter Mueller, Yuan Ji</i>	
<b>A Nonparametric Bayesian Model for Local Clustering</b> .....	29
<i>Juhee Lee</i>	
<b>Testing Goodness of Fit of Protein Conformational Sampling Models</b> .....	29
<i>Mehdi Maadooliat</i>	
<b>Improving Compton Scatter Camera Image Resolution Using Classification and the Metropolis-Hastings Algorithm</b> .....	30
<i>Dennis Mackin</i>	

**BANQUET ADDRESS**

<b>Repetition and Surprise, Rehearsal and Reinvention</b> .....	30
<i>Mark Hansen</i>	

## **MODELING, ANALYZING, AND VISUALIZING HIGH-DIMENSIONAL DATA IN GENOMICS**

<b>Fast Graphical Model Estimation and Its Applications</b> .....	31
<i>Daniela Witten</i>	
<b>Conditional Network Testing in High-dimensional Dependent Data</b> .....	32
<i>Gary Gadbury</i>	
<b>Robust Identification of Conditional Gene Expression in Development of Onthophagus Beetles</b> .....	32
<i>Guilherme V Rocha</i>	

## **VISUALIZATION AND COMPUTATIONAL METHODS FOR ACTIGRAPHY DATA**

<b>Powerful Actigraphy Data Through Functional Representation</b> .....	33
<i>Jimin Ding</i>	
<b>Reliability and Reproducibility Issues in Accelerometer-Based Estimates of Physical Activity</b> .....	33
<i>Julia Kozlitina</i>	
<b>Movelets: A Dictionary of Movement</b> .....	33
<i>Jeff Goldsmith</i>	

## **DEVELOPING INTELLIGENCE IN UNMANNED GROUND SYSTEMS**

<b>Some Thoughts on Experimentation Philosophy in the Robotics CTA</b> .....	34
<i>Barry Bodt</i>	
<b>Preliminary Performance Evaluation of Autonomous Mobility in Small UGVs</b> .....	34
<i>Alberto Lacaze</i>	
<b>Using Expectations to Drive Cognitive Behavior</b> .....	35
<i>Unmesh Kurup</i>	

## **COMPUTATIONAL TOOLS AND STATISTICAL METHODS WITH MEDICAL APPLICATIONS**

<b>Graph-Based Signal Integration for High-Throughput Phenotyping</b> .....	35
<i>Jorge Herskovic, Elmer Bernstam</i>	
<b>Extending the Grammar of Graphics for Genomic Data: an R Implementation</b> .....	36
<i>Tengfei Yin, Dianne Cook, Michael Lawrence</i>	
<b>Massive Parallelization of Serial Inference Algorithms for a Complex Generalized Linear Model</b> .....	36
<i>Marc Suchard, Shawn Simpson, Ivan Zorych, Patrick Ryan, David Madigan</i>	

## **APPLICATIONS OF INTERACTIVE GRAPHICS IN R**

<b>Using R and Web Technologies to Create Analytics and Apps: Part of Making Statistics Relevant in a Large Organization</b> .....	37
<i>Chad Shaw</i>	
<b>Exploring Statistical Strategies for Use in Challenge-Response Experiments</b> .....	37
<i>Matthew S. Shotwell, Kenneth J. Drake, Veniamin Y. Sidorov, John P. Wikswo</i>	
<b>The Anscombe Data Sets: Explained and Expanded</b> .....	38
<i>Jurgen Symanzik</i>	

## **GENERAL PARALLEL COORDINATES**

<b>Cluster Detection and Visualization with Generalized Parallel Coordinates</b> .....	38
<i>Rida Moustafa</i>	
<b>Visual Cluster and Outlier Detection with L-plot</b> .....	39
<i>Michael D. Larsen, Rida E. Moustafa, Ali S. Hadi,</i>	
<b>Visual Analytics Approach for Social Network Interaction</b> .....	39
<i>Jie Cong, Rida Moustafa</i>	

**CONTRIBUTED PAPER SESSION IV**

**Parallel Monte Carlo Simulation for the Sensitivity Analysis of Expected Shortfall by Means of a Second-Order Approximation** ..... 39  
*Güven Gul Polat*

**Comparison of Binary Discrimination Methods for High Dimension Low Sample Size Data** ..... 40  
*Ady Bolívar-Cime, J.S. Marron*

**Factor Model for Forecasting with Multi-collinearity and Nonlinear Dependence** ..... 40  
*Joseph Egbulefu*

**Relations Between Attentional Structure and Attentional Function: Utilization of Alternative Statistical Approaches** ..... 40  
*Paulina Kulesz*

**Author Index**