

35th Symposium on the Interface: Computing Science and Statistics 2003

Security and Infrastructure Protection

Computing Science and Statistics Volume 35

**Salt Lake City, Utah, USA
12-15 March 2003**

ISBN: 978-1-61567-069-7

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© (2003) by the Interface Foundation of North America
All rights reserved.

Printed by Curran Associates, Inc. (2009)

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

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
Web: www.proceedings.com

Table of Contents

Keynote Address

Considerations of Inspections for Homeland Security with Cross Linkages to Quality Control, Game Theory, and Stochastic Simulation (Keynote Address Abstract) James R. Thompson.....	1
Models of Homeland Security with Borrowings from SPC and Game Theory (Keynote Paper), James R. Thompson.....	2

Short Courses

Statistical Methods in Computer Security (Abstract and Presentation) David J. Marchette.....	17
Topics in Computational Statistics with MATLAB (Abstract and Presentation) Jeffrey L. Solka.....	18

R

Technology Transfer in Industry using an R-COM Interface (Abstract) Scott Wander Wiel, John Chambers, Suresh Goyal, David James.....	20
Inter-System Interfaces for S (Abstract) Duncan Temple Lang.....	21
Sparse Linear Algebra for the R Language (Abstract and Presentation) Pin Ng, Roger Koenker.....	22

Bioinformatics

Exploratory Detection of Differential Gene Expression (Abstract) John D. Storey.....	24
Statistical Issues in the Design of Microarray Experiments (Abstract) Jean Yee Hwa Yang, Terry Speed.....	25
Exploration, Normalization, Summaries and Software of Affymetrix GeneChip Probe Level Data (Abstract and Presentation), Rafael Irizarry.....	26

Security and Infrastructure Protection: Interface of Computing Science and Statistics to the Rescue

Statistical Opportunities in Network Security (Paper and Presentation) David J. Marchette.....	28
Using Statistics to Detect and Thwart Denial of Service Attacks (Abstract) Carla Brodley.....	39
Maximizing Quality and Value in Security: Challenges for Computer Scientists, Statisticians and Their Clients (Abstract) , Arnold F. Goodman.....	40

Usability and Accessibility of Visualization Tools for Health Statistics

Usability Testing of Map Designs (Paper) Linda William Pickle.....	42
Accessible Graphics on the World Wide Web (Extended Abstract and Presentation) Dan J. Grauman.....	57
Java-based Dynamic Linked Micromap Plots (Abstract and Presentation) Jim X. Chen, Xusheng Wang, Daniel B. Carr, B. Sue Bell, and Linda W. Pickle.....	59
Web Design and Usability Guidelines: An Evidence-based Approach (Abstract) Sanjay Koyani.....	60

Statistical Graphs

Graph-Theoretic Latent Class Discovery and its Robustness to Minimal Discriminating Set Choice (Abstract and Presentation) Jeff Solka, Carey Priebe, David Marchette.....	62
Asymptotic Theory for the Domination Number of Random Class Cover Catch Digraphs (Paper) , John C. Wierman and Pengfei Xiang.....	63
Bayesian Models for Sparse Edge Weighted Directed Graphs (Abstract and Presentation) Deepak Agarwal.....	75
Sensitivity Analysis of Graph Distance Measures (Paper) Julian Sorenson, Peter Dickinson and Martina Schubert.....	76
Latent Variable Models for Link Analysis of Similarity Data (Paper) Juan K. Lin.....	89

Financial Risk and Fraud Detection

Audit at a Crossroads (Paper) Conan C. Albrecht.....	103
Credit Scoring Using Bureau and Other Data (Abstract) Garry K. Ottosen.....	112
Computational Statistics with Spreadsheets Towards Efficiency, Reproducibility and Security (Paper) , G. Aydin, W. Härdle, E. Neuwirth.....	113
Some Loans are More Equal than Others: Third-Party Originations and Defaults in the Subprime Mortgage Industry (Abstract and Presentation) Scott D. Grimshaw, Grant R. McQueen, William P. Alexander, Barrett A. Slade.....	128

Computationally Challenging Statistical Methods in Genetics

Enumeration and Simulation of Marriage Node Graphs on Zero Loop Pedigrees (Abstract) Alun Thomas and Chris Cannings.....	130
Statistical and Computational Issues in Mapping Genes in Animal Populations (Abstract) Fengxing Du.....	131
Computations in Animal Breeding (Abstract and Presentation) Ignacy Misztal and Romdhane Rekaya.....	132

Statistical Analysis and Probabilistic Modeling of Internet Traffic

Statistical Clustering of Internet Communication Patterns (Abstract and Presentation) Felix Hernandez-Campos, A. B. Nobel, F. D. Smith, K. Jeffay.....	134
A Solution to the Bandwidth Allocation Problem for the Internet (Abstract) William S. Cleveland, Jin Cao and Don X. Sun.....	135
The Joint Distribution of Internet Flow Sizes and Durations (Abstract and Presentation) Cheolwoo Park and J S. Marron.....	136

Data Analysis and Visualization of Graph Data

Using Graphs to Explore Communication Networks (Abstract) Chris Volinsky.....	138
Graphs and EDA in Computational Biology (Abstract) Robert Gentleman.....	139

Visual Exploration of Graph Data (Abstract and Presentation) Deborah F. Swayne, Duncan Temple Lang and Andreas Buja.....	140
---	-----

Computational Methods

Cost-Sensitive Classifier Selection Using the ROC Convex Hull Method (Paper and Presentation), Ross Bettinger.....	142
--	-----

A New Imagery Classification Methods Using Spatial Covariance Information (Paper) James A. Shine and Daniel B. Carr.....	154
---	-----

A Risk-Utility Framework for Data Swapping (Abstract and Presentation) Shanti Gomatam, Alan Karr and Ashish Sanil.....	159
---	-----

Optimal Stopping of a Risk Process with Interest Rates (Paper and Presentation) Bogdan Krzysztof Muciek.....	160
---	-----

A Likelihood Approach for Determining Cluster Number (Paper) William D. Shannon, Tsvika Klein and Robert Culverhouse.....	166
--	-----

Advances in Face/Pattern Recognition

Obtaining Smooth Directional Field Estimates for Fingerprint Images (Abstract) Sarat C. Dass.....	176
--	-----

Fast Face Detection with a Boosted CCCD Classifier (Paper), Diego A. Socolinsky, Joshua D. Neuheisel, Carey E. Priebe, Jason DeVinney, David Marchette.....	177
--	-----

Computer-Intensive Statistical Methods

Parametric Bootstrap Confidence Intervals in Small Area Estimation Problems (Abstract) Snigdhanu Chatterjee and P. Lahiri.....	186
---	-----

Nonparametric Tail Index Estimation (Abstract) Tucker McElroy and Dimitris Politis.....	187
--	-----

Bayesian Inference in Single-Layer Neural Networks (Abstract) Robert Paige.....	188
--	-----

Selecting Optimal Block Lengths for Block Bootstrap Methods (Paper) S. N. Lahiri.....	189
--	-----

Computational Statistics and NSA

Statistical Analysis of Massive Data Streams: Overview of a CATS Workshop (Abstract)
Sallie Keller-McNulty.....204

A-Family Priors: Smoothing Multinomial Data (Abstract)
Jeff Benedict.....205

Model Building and Diagnostics for Massive Data Sets (Abstract)
David W. Scott.....206

Graphics and Visualization

Visual Data Mining for Quantized Spatial Data (Abstract and Presentation)
Amy Braverman.....208

Interactive Federal Statistical Data on the Web Using “NVIZN” (Paper and Presentation)
Jon Hurst, Jürgen Symanzik and Lacey Gunter.....209

Visualizing Random Forests (Abstract)
Adele Cutler and Leo Breiman.....222

Interactive Spinograms (Abstract)
Heike Hofmann and Martin Theus.....223

Scatterplots for Massive Datasets (Abstract and Presentation)
Martin Theus, Di Cook and Heike Hofmann.....224

Sensors for Biological Threats

Identification of Bio-warfare Agents and Other Applications of Molecular Biology
(Abstract), Todd Ritter.....226

Rapid Microbe Detection with Fluidized Bed Capture and Concentration (Abstract)
Bart Weimer and Marie Walsh.....227

Early Diagnosis of Biological Threats: Progress and Challenges (Abstract)
Stephen S. Morse.....228

Best of KDD-2002

Customer Lifetime Value Modeling and its Use for Retention Planning (Abstract)
Saharon Rosset, Elnat Neumann, Uri Eick, Nurit Vatnik.....230

Doing Something Useless Slightly Faster: The State of the Art in Time Series Data Mining (Abstract), Eamonn Keogh.....	231
--	-----

Homeland Security and Related Issues

Areas of Homeland Security: At the Computational Statistical Interface (Abstract) Deborah Leishman.....	233
---	-----

Rule-Based Anomaly Pattern Detection for Detecting Disease Outbreaks (Abstract) Andrew Moore.....	234
---	-----

Pointers from Research on Data Confidentiality and Data Quality (Abstract and Presentation), Ashish Sanil.....	235
--	-----

Modern Text Processing and Distribution

The Journal of Statistical Software (Abstract and Presentation) Jan de Leeuw.....	237
---	-----

Preparing Electronic Books (Abstract and Presentation) Edward J. Wegman and Amy Braverman.....	238
--	-----

Electronic Books for Experts and Users (Paper and Presentation) Zdenek Hlávka.....	239
--	-----

Environmental Statistics

A Spatial Model for Chronic Wasting Disease in Rocky Mountain Mule Deer (Paper) Christopher H. Mehl and Craig J. Johns.....	251
---	-----

Spatial Statistics in the Presence of Location Error (Abstract and Presentation) John Kornak and Noel Cressie.....	262
--	-----

Separating Signal from Noise in Global Warming (Paper) Bert W. Rust.....	263
--	-----

Predictive Mapping of Forest Characteristics for Fire Risk Assessment (Abstract) Gretchen Moisen, Tracey Frescino, Cheng Huang, Jim Vogelmann, Zhiliang Zhu.....	278
--	-----

Nonparametric Modeling of Soil Characteristics of Crop Models (Abstract) Stephan R. Sain and Doug Nychka.....	279
---	-----

Social Networks and Statistics

Random-Effects Models for Network Dependence (Abstract) Peter Hoff.....	281
Ultra-Robust and Scalable Networks Based on Hierarchies (Abstract and Presentation) Peter Dodds, Duncan Watts and Charles Sabel.....	282
Statistical Models, Degeneracy and Inference for Social Networks (Abstract) Mark S. Handcock.....	283

Best of the International Association for Statistical Computing

Incremental Tree-Based Missing Data Imputation with Lexicographic Ordering (Paper and Presentation) , Claudi Conversano and Roberta Siciliano.....	285
Many Faces of a Tree (Paper) Simon Urbanek.....	306
A Wildlife Simulation Package (<i>WiSP</i>) (Paper and Presentation) Walter Zucchini, David Borchers, Stefan Kirchfeld and Martin Erdelmeier.....	315

Digital Government Research in Support of Federal Statistics

Using Ontology as Generalized Metadata Schema for Access to Distributed Heterogeneous Data Sources (Abstract) , Edward Hovy.....	330
Interfaces to a Statistical Knowledge Network (Abstract) Gary Marchionini.....	331
New Approaches to Mobile Computing for Field Data Collection (Abstract) Sarah Nusser.....	332

Smoothing and Nonparametric Feature Detection

A SiZer Analysis of IP Flow Start Times (Abstract) J. S. Marron, Felix Hernandez-Campos and F. D. Smith.....	334
Longitudinal Kernel Regression (Abstract) Naisyin Wang, Raymond J. Carroll, Xihong Lin and Ziding Feng.....	335
Semiparametric Regression Smoothing and Feature Detection in Time Series (Paper) Michael G. Schimek.....	336

Safety and Security

- [Traffic Safety Analysis: A Data Mining Approach \(Paper\)](#)
J. Michael Hardin, Michael Conerly and Wade Watkins.....353
- [Bayesian Inductively Learned Modules for Safety Critical Systems \(Paper\)](#)
Jonathan E. Fieldsend, Trevor C. Bailey, Richard M. Everson, Wojtek J. Krzanowski,
Derek Partridge and Vitaly Schetini.....367
- [Waypoint Analysis for Command and Control \(Abstract\)](#)
Mark Irwin, David Wendt and Noel Cressie.....381
- [Continually Improving Stream Analysis for Network Security \(Paper and Presentation\)](#)
Nancy J. McMillan, Douglas D. Mooney and David A. Burgoon.....382
- [A Micro-Scale Epidemiological Simulation for Management of Disease \(Abstract\)](#)
Sid Baccam, Stephen Eubank and Catherine Macken.....397

Public Health Preparedness and Response in Crisis

- [Using Design-Based Adaptive Sampling Procedures in Site Decontamination \(Paper\)](#)
Myron J. Katzoff, Abera Wouhib and Joe Fred Gonzalez, Jr.....399
- [Game Theory and Risk Analysis for the Smallpox Threat \(Abstract and Presentation\)](#)
David Banks.....409

Best of the *Journal of Computational and Graphical Statistics*

- [Penalized Survival Models and Frailty \(Abstract\)](#)
V. Shane Pankratz, Patricia M. Grambsch and Terry M. Therneau.....411
- [Adaptive Order Selection for Spline Smoothing \(Abstract\)](#)
Randy Eubank, Chunfeng Huang and Suojin Wang.....412
- [An Adaptive Spatial Scan Density Estimation Method \(Abstract\)](#)
Ramani S. Pilla, Peng Tao and Carey Priebe.....413

Infrastructure Security

- [Energy Infrastructure Vulnerability Assessments \(Abstract\)](#)
Jeff Dagle.....415
- [PNNL and International Border Security \(Abstract\)](#)
William C. Cliff.....416

Electricity Infrastructure Security (Abstract) Thomas Kropp.....	417
---	-----

Statistical Computing (Refereed Papers)

RGL: A R-Library for 3D Visualization with OpenGL (Paper and Presentation) Daniel Adler, Oleg Nenadic and Walter Zucchini.....	419
---	-----

Constructive Ensembles for Time Series in Econometrics and Finance (Paper) H. D. Vinod.....	430
--	-----

The Quickest Sequential Detection of Intrusions in Computer Networks (Abstract) Boris Rozoskii, Rudolf Blazek, Hongjoong Kim and Alexander Tartakovsky.....	448
--	-----

Implementing Legacy Statistical Algorithms in a Spreadsheet Environment (Paper and Presentation), Stephen W. Liddle and John S. Lawson.....	449
---	-----

Bioinformatics

A Bayesian Mixture Model for Bayesian Gene Expression (Abstract and Presentation) Kim-Anh Do, Peter Mueller and Feng Tang.....	472
---	-----

A Simple Approach to Accommodating Interactive and Batch Processes on a Bioinformatics Cluster (Abstract) Warren M. Snelling, John W. Keele and Gregory P. Harhay.....	473
---	-----

Selecting an Optimal Rejection Region for Multiple Testing (Paper) David R. Bickel.....	474
--	-----

Statistical Methods for Spot Detection with Macroarray Data (Paper) Yi Xie, Adele Cutler, Bart Weimer and Andrejus Parfionovas.....	490
--	-----

Statistical Issues in Computer Security

Email Worm Propagation on Random Graphs (Paper) Stephan Bohacek.....	507
---	-----

User Profiling for Intrusion Detection in Windows NT (Paper) Tom Goldring.....	524
---	-----

A Stochastic Model of Computer Intrusions for Evaluations and Exercises (Abstract and Presentation), Robert P. Goldman.....	528
---	-----

Multi-Level Monitoring and Fuzzy Clustering to Detect Cyber Attacks (Abstract) Dipankar Dasgupta, Jonatan Gomez and Fabio Gonzalez.....	529
--	-----

Interactive GeoGraphics for the Web

Integrated Climate Database (Abstract) Dan Dansereau and Robert R. Gillies.....	531
Web Cartography for Municipal Government: An Accessibility Case Study (Abstract) Robert Edsall.....	532

Design and Statistical Computation

Application of Simulated Annealing to D-optimal Design for Polynomial Regression with Correlated Observations (Abstract), Zewen Zhu and Daniel C. Coster.....	534
Experimental Design for Body Image Testing (Paper) Craig J. Johns and Russel L. Boice.....	535
A Stabilized Lugannani-Rice Formula (Paper) George R. Terrell.....	546
Computation of the Normalization Constant for Exponentially Weighted Dirichlet Distribution Integrals (Paper), Alan Genz and Paul Joyce.....	557
A Plot for Visualizing Multivariate Data (Paper and Presentation) Rida E. A. Moustafa.....	564

Nonparametrics (Refereed Papers)

Estimating Partially Linear Models Using Wavelets: A Nonlinear Backfitting Algorithm (Paper), Leming Qu.....	577
Nonlinear Smoothers in Two Dimensions for Environmental Data (Paper) Karen Kafadar and Max D. Morris.....	586
A Comparison of Filters and Wrappers Methods for Feature Selection Methods in Supervised Classification (Paper), Edgar Acuña.....	613
Novel Methods for Multivariate Ordinal Data Applied to Olympic Medals, Risk Profiles, Genomic Pathways, Genetic Haplotypes, Pattern Similarity, and Array Normalization (Paper), Knut M. Wittkowski.....	626

Data Management for Statistical Databases

Database Technology for Statistical Data (Abstract) Arie Shoshani.....	647
---	-----

Metadata Usage in Statistical Computing (Paper)	
Wilfried Grossmann.....	648

Data Structures for HIV and AIDS Notification and Analysis (Abstract and Presentation)	
Andrew Westlake.....	664

Surveillance

Multiscale ‘Spatial’ Analysis of Network Data: Putting Wavelets on Graphs (Abstract)	
Eric D. Kolaczyk.....	666

Social Networks and Computer Networks (Abstract and Presentation)	
John Rigsby and Jeff Solka.....	667

Simultaneous Selection of Features and Metric for Optimal Nearest Neighbor Classification (Paper and Presentation)	
David A. Johannsen, Edward J. Wegman, Jeffrey L. Solka and Carey E. Priebe.....	668

Prediction of Catastrophic Events

Detecting Features in Seismic and Geodetic Data (Abstract)	
Andrea Donnellan, Robert Granat and John Rundle.....	685

Predicting Damaging Climate Events: Methods, Examples, and Public Reaction (Abstract), David W. Pierce and Tim P. Barnett.....	686
--	-----

Predicting and Comprehending Asteroid Impacts (Abstract and Presentation)	
Clark R. Chapman.....	687

Graphics for Bio and Chem Informatics

Graphic-Centric, Computationally-Efficient Recursive Partitioning (Abstract)	
James Vivian, S. Stan Young and Christophe Lambert.....	689

Applications of Computational Geometry, Statistical Analysis and Graphics to the Study of Molecular Systems (Abstract), Daniel B. Carr and Iosif Vaisman.....	690
---	-----

Statistical Methods to Compress and Query Large Databases and Data Streams

Data Stream Algorithmics (Abstract)	
S. Muthukrishnan.....	692

Efficient Processing of Massive Data Streams for Mining and Monitoring (Abstract and Presentation), Mirek Riedwald, Johannes Gehrke, Alan Demers, Abhinandan Das, Alin Dobra.....693

Wavelets for Efficient Querying of Large Multidimensional Data Sets (Abstract and Presentation), Cyrus Shahabi.....694

Data Mining Combat Simulations

Data Mining Combat Simulations: An Emerging Opportunity (Abstract and Presentation) Barry A. Bodt.....696

Regression Tree Analysis of Battle Simulation Data (Abstract) Wei-Yin Loh.....697

Robust Modeling Based on L_2E Applied to Combat Simulation Data (Paper) David B. Kim.....698

Discovery of Battle States Knowledge from Multi-Dimensional Time Series Data (Abstract) T. W. Liao, B. Bodt, J. Forester, C. Hansen, E. Heilman, C. Kaste, J. O'May.....708

Forests at Risk

Satellite to the Public in Near Real-Time: Providing Active Wildfire Information with MODIS Rapid Response (Abstract) Mark Finco, Brad Quayle, Rob Sohlberg and Jacques Descloitres.....710

Identifying "Redtops": Classification of Satellite Imagery for Tracking Mountain Pine Beetle Progression through a Pine Forest (Paper) Richard Cutler, Leslie Brown, James Powell, Barbara Bentz and Adele Cutler.....711

Design Attributes for Sampling Rare Ecological Events in Forest Ecosystems: Lichens in the Pacific Northwest (Abstract), Thomas C. Edwards and Richard Cutler.....729

Nonparametric Methods and Applications

Multivariate Density Estimation with Permuted Variable-Values (Abstract and Presentation), Sridevi Parise, Padhraic Smyth and Sergey Kirshner.....731

Computational Challenges in Computing Nearest Neighbor Estimates of Entropy for Large Molecules (Paper and Presentation) E. James Harner, Harshinder Singh, Shenggiao Li and Jun Tan.....732

Mixture Transitions for Edge Preservation in Kalman Filtering (Abstract) Mark Fitzgerald.....743

Statistical Learning Theory and Statistics: Embracing New Technologies (Abstract)
Kevin Watanabe.....744

Using a LOESS Smoother to Estimate the Parameters of an Angular Dependent
Distribution of HRR Data (Abstract)
Bradley C. Wallet, Robert W. Hawley and Troy L. Klein.....745

**The following paper was inadvertently omitted from the 2001
Proceedings Volume 33**

Interval Computation of Gamma Probabilities and their Inverses (Paper)
Trong Wu.....747

Author and Organizer Index