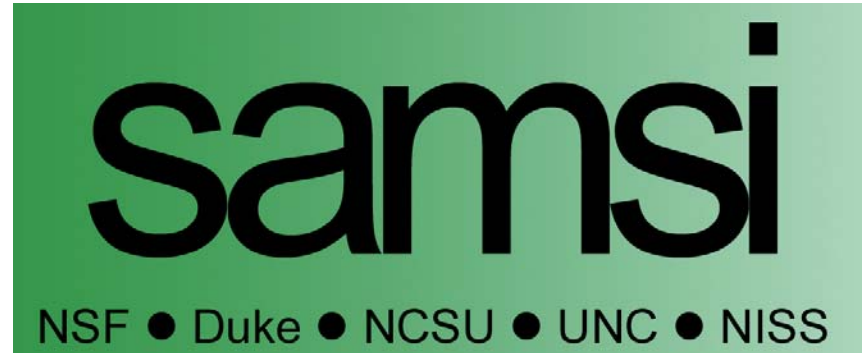


NISS



NISS Affiliates

NISS/SAMSI University Affiliates

JSM Meeting
San Francisco
August 3, 2003

Agenda

- Welcome and Introductions
- 2002-03 in Review
- Current Program Status
- Plans for 2003-04
- SAMSI: 2003-04 and Beyond
- 2003 Affiliates Survey
- Initiatives
- Adjourn

SAMSI Grand Opening



Affiliates Technology Days



Affiliates Short Course



2003 Planning Meeting



Current NISS Affiliates

- **Corporations:** Amgen, Avaya Labs, *Aventis Pharmaceuticals*, General Motors, GlaxoSmithKline, Merck & Company, MetaMetrics, Pfizer Inc., RTI International, SAS Institute, SPSS, Telcordia Technologies
- **Federal Agencies and National Laboratories:** Bureau of Labor Statistics, Bureau of Transportation Statistics, Census Bureau, Los Alamos National Laboratory, National Agricultural Statistics Service, National Center for Education Statistics, National Center for Health Statistics, National Institute of Standards and Technology, National Security Agency, Pacific Northwest National Laboratory

Current NISS/SAMSI University Affiliates

- American (Mathematics & Statistics), *UCLA* (Statistics/Statistical Consulting Center), Carnegie Mellon (Statistics), Duke (Statistics and Decision Sciences/Mathematics), Emory (Biostatistics), *Florida State* (Statistics), *Georgia* (Statistics), *Illinois Urbana-Champaign* (Statistics), Iowa (Statistics), Iowa State (Statistics), Johns Hopkins (Mathematical Sciences), Maryland College Park (Mathematics), Maryland Baltimore County (Mathematics and Statistics), Michigan (Statistics/Biostatistics), *Missouri Columbia* (Statistics), North Carolina State (Mathematics), North Carolina State (Statistics), North Carolina Chapel Hill (Biostatistics), North Carolina Chapel Hill (Mathematics), North Carolina Chapel Hill (Statistics), *Oakland* (Mathematics & Statistics), Ohio State (Statistics), Pennsylvania State (Statistics), Purdue (Statistics), *Rice* (Statistics), Rutgers (Statistics), Southern Methodist (Statistical Sciences), Stanford (Statistics), Texas A&M (Statistics)

Principal Current Activities - 1

- Research
 - Affiliates involved directly: GM, GSK, BLS, BTS, Census, LANL, NCES, CMU, Duke, Iowa, NCSU, UNC, SMU
 - SAMSI Programs
 - Affiliates Proposal Development Fund
- Education & Information Dissemination
 - Technology Days (3 in 2002-03)
 - Short Courses (2 in 2002-03)
 - SAMSI Workshops

Principal Current Activities - 2

- Professional Development
 - Federal agency affiliate postdocs
 - SAMSI research visits
- Networking and Affiliate Support
 - 2 general meetings/year
 - AJLS
 - NAIP
 - APDF

Recent Events at NISS That Can Impact Affiliates

- Burgeoning postdoc program
 - 10 this year
- Emerging research program in bioinformatics
 - NISS-CIIT Joint Postdoctoral Fellowship: high-density microarrays
 - HDF Project: drug discovery to cure Huntington's disease
- HSD proposal: Statistical Disclosure Limitation for Geospatial Image Data
 - Includes workshop
- ACT proposal: Workshop on Mathematical Sciences Research to Meet Counterterrorism Needs

Program Goals for 2003-04

- Continue to grow and diversify
 - Add at least 5 corporate affiliates
 - Add at least 10 NISS/SAMSI University Affiliates
- Expand and improve current activities
- Increase co-sponsorship of events, enhancing ARA value
 - EURANDOM alliance
 - MMR 2004 (June 2004, Santa Fe)
- Explore several new initiatives

2003-04 Technology Days

- Data Confidentiality: Fall 2003, DC
- SDL for Geospatial Image Data: January 2004, NISS
- Overarching Issues in Risk Assessment: Spring 2004, Iowa State
- Implications of NSF Cyberstructure Initiative for Statistics: TBD
- Co-sponsorship of SAMSI DM&ML mini-workshops

samsi

NSF • Duke • NCSU • UNC • NISS

Location: NISS building
Research Triangle Park
North Carolina

Information: www.samsi.info

2002-03 Highlights

- Grand Opening Ceremony
- Full slate of activities
 - Three programs
 - 7 program workshops
 - Outreach: 3 undergrad workshops, IMSMW '03
 - 6 SAMSI Distinguished Lectures
- Full complement of people
 - 7 postdocs
 - 3 SAMSI-University Fellows
 - 40+ research visitors
 - 700+ workshop participants

SAMSI Programs

- 2-3 research programs per year
- At interfaces: Involve both statistics and applied math (also probability, computational sciences, operations research)
- Framed and guided by disciplinary needs
- “catalytic rather than conclusional”
- Multiple types: Focused Study Programs, Synthesis Programs, Pilot Programs

SAMSI Programs for 2003-04

- **Network Modeling for the Internet** (Fall, 2003)
- **Data Mining and Machine Learning** (Entire year)
- **Multiscale Modeling and Control Design** (Spring, 2004)

Network Modeling for the Internet

(Fall, 2003)

- *High-Level Goal*: Stimulate interaction among applied probabilists studying heavy traffic queueing theory and fluid flow models, internet traffic measurers/modelers and hardware/software architects, and statisticians.
- *Scientific Committee*: K. Jeffay (UNC), J. Landwehr (Avaya), J. P. Lehoczky (CMU), J. S. Marron (UNC/SAMSI), R. Williams (UCSD), W. Willinger (AT&T), D. Towsley (UMass)
- *Workshops*
 - NISS Affiliates Internet Tomography Technology Day: March 29, 2003
 - Measurement, Modeling, Heavy Traffic: Sept. 17-20, 2003
 - Internet Tomography, Sensor Networks: Oct. 13-15, 2003

Multiscale Modeling and Control Design (Spring, 2004)

- *High-Level Goal*: Forum to investigate
 - Nano, micro and mesoscale modeling methodologies
 - Techniques for real-time implementation of system models
 - Robust control designs for advanced material architectures
- *Scientific Committee*: G. Forest (UNC), D. Cioranescu (Universite Pierre et Marie Curie), A. Gelfand (Duke), D. Schaeffer (Duke), M. Salapaka (Iowa State), R. Smith (NCSU), C. Wickle (Missouri), M. Wright (NYU)
- *Workshops*
 - Opening Workshop: January 17-21, 2004
 - Soft Matter and Nano-Materials: February 15-17, 2004
 - Granular Flow: April 15-17, 2004

DMML: Motivation

- DMML: discovery of patterns, information and knowledge in large, complex (and, often, unstructured) data sets
- Proliferation of techniques over the past several years
- Incomplete understanding of fundamental statistical and computational issues
- Affiliate interest

Goals of the Program

- Significantly advance understanding of statistical and computational issues
- Articulate future research needs, especially from the perspective of the statistical sciences
- Catalyze formation of collaborations among statistical, mathematical and computer scientists to pursue the research agenda

Potential Foci

- Large p , small n inference
- Bioinformatics
- Support vector machines
- Computational experiments that relate performance of DMML techniques to problem characteristics
- Novel, complex forms of data, especially text
- Statistical questions inherent in DMML techniques: sampling, model selection and search, robustness, data quality, multiplicity
- Computational scalability

Participants - 1

- **Scientific Committee:** *David Banks* (Duke; Co-chair), *Mary Ellen Bock* (Purdue), *Jerome Friedman* (Stanford), *Alan F. Karr* (NISS; Chair), *David Madigan* (Rutgers), *William DuMouchel* (AT&T), *Warren Sarle* (SAS Institute)
- **SAMSI-University Fellow:** *Bernard Clarke* (UBC)
- **Faculty Fellows:** *David Banks* (Duke), *Marc Genton* (NCSU), *Feng Liang* (Duke), *Jacqueline Hughes-Oliver* (NCSU)
- **Postdocs:** *Ernest Fokoue*, *Xiaodong Lin* (joint with NISS)
- **Graduate Students:** 4 from Duke, NCSU

Participants - 2

- **Research Visitors:** Susie Bayarri (Valencia), Hugh Chipman (Waterloo), *Di Cook* (Iowa State), William DuMouchel (AT&T), *Prem Goel* (Ohio State), Karen Kafadar (Colorado Denver), *David Madigan* (Rutgers), *Marianthi Markatou* (Columbia), *Vijay Nair* (Michigan), Jesus Palomo (Madrid), *Dongchu Sun* (Missouri)
- **NISS Staff:** Gary McDonald, Alan Karr, Ashish Sanil, Stanley Young
- **NISS Postdocs:** Murali Haran, Jack Liu
- **Other Triangle Area Researchers:** Alex Tropsha (UNC), Wei Wang (UNC)
- ...

Program Activities

- Tutorials/Opening Workshop: September 6-10, 2003
- Closing Workshop: May 20-22, 2004
- Mini-workshops (1 day)
 - Explore ramifications of DMML for other issues
 - Treat related problems such as text mining
 - “Mid-term” workshop in January, 2004
- Research
 - Organized into working groups
 - Focused on testbed data sets
- Monograph detailing conclusions and recommendations
 - To be published in a new SAMSI subseries of the ASA-SIAM Series on Statistics and Applied Probability.

Kickoff Workshop

- September 6-10, 2003
- **Invited Speakers:** Leo Breiman (UC Berkeley), *Di Cook* (Iowa State), William DuMouchel (AT&T), Michael Jordan (UC Berkeley), *David Madigan* (Rutgers), J. S. Marron (UNC & SAMSI), Robert McCulloch (Chicago), *Jeff Schneider* (Carnegie Mellon), Peter Westfall (Texas Tech)
- **Activities:** Tutorials (9/6), Birds-of-a-Feather Session, Poster Session, Poster Sales Talks, Second Chance Seminar, Young Researchers Session, Working Group Meetings (9/10)

Tentative Programs for 2004-2005

- **Computational Modeling of Infectious Disease:** (Full Year, 04-05): Contacts - *Thomas B. Kepler* (Duke) and *Denise Kirschner* (Michigan)
 - Population dynamics
 - epidemiology, social network theory, immune-response modeling
 - microbial ecology, host-pathogen co-evolution, drug resistance, evolution
 - Genomic dynamics
 - comparative genomics, molecular evolution, proteomics
 - gene expression modeling and microarray data analysis
 - drug target identification and vaccine design

- **Social Sciences:** (Full year, 04-05). Contact – *Alan Karr* (NISS); possible emphases include
 - agent-based simulation
 - social networks
 - genetic and environmental influences on behavior
 - longitudinal analysis
 - causality
- **Data Assimilation:** (Spring, 05): Contacts - *Chris Jones* (UNC-CH), *K. Ide* (UCLA), *D. Nychka* (NCAR); possible emphases include
 - how to manage large data-sets
 - improving models, i.e. can models learn?
 - assimilation of non-prognostic variables
 - optimal design of experiments based on the results of data assimilation
 - assimilation of data corresponding to subsidiary variables in coupled models, such as atmospheric data in coupled atmosphere-ocean models, or Lagrangian data in ocean models

Potential Future Programs

- 2005-06
 - High Dimensional Inference and Random Matrices
 - Financial Mathematics
 - ??
- Hot Topic Workshops
 - not necessarily related to major programs
 - short lead times (< 1 year) from proposal to workshop

Opportunities for Senior Researchers

- Program proposals and leadership
- Research visits (short- and long-term)
 - SAMSI provides expenses, space, support
- Workshops
 - SAMSI or ARA provides expenses
- SAMSI-University Fellowships (one semester or full year)
 - SAMSI and university provide partial salary

SAMSI Proposal Process

- For year n programs,
 - Pre-proposals due September 15 of year n-2
 - Full proposals due March of year n-1
 - Decisions, including budget, made April of year n-1
 - Urgent programs can be fast-tracked
- Ideas and informal discussions welcome at any time
 - Directorate
 - National Advisory Council (Co-chairs: Peter Bickel, Margaret Wright)

Proposal Evaluation Criteria

- Impact on statistical and mathematical sciences and their interface
- Qualifications and commitment of program leaders
- Breadth of development of human resources, including the extent of diversity
- Impact on cross-disciplinary research
 - Span of the scientific problems covered
 - Likelihood of significant technology transfer
- Effects on education
- Timeliness
- Availability of leveraging funds

Opportunities for Young Researchers

- Postdoctoral fellowships
 - Typically 2 years—one at SAMSI and one at NISS, CRSC, university, ...
 - SAMSI provides salary, space, support, travel
- Research visits (short- and long-term)
 - Some positions with partial salary support
 - Numerous positions with expense support
- Workshops
 - SAMSI or ARA provides expenses

Opportunities for Graduate Students

- Research Programs
 - Through Duke, NCSU, UNC: SAMSI provides stipend, space, support; university provides tuition
 - Visit with advisor: SAMSI provides expenses, space, support
- Program Workshops
 - SAMSI or ARA provides expenses
- IMSMW
 - 9 days: end of July
 - 6 teams work on “industrial” projects presented by experienced scientists and engineers: 6 students, 1 problem owner, 1-2 faculty mentors
 - SAMSI provides expenses

Opportunities for Undergraduates

- Two-day Undergraduate Workshops on weekends during the academic year
 - For juniors and seniors
 - Description of SAMSI programs, and hands-on experience with one of them
 - SAMSI provides expenses
- SAMSI/CRSC Undergraduate Workshop for one week in early June
 - For juniors, seniors, and high-school teachers
 - Description of SAMSI programs and intensive hands-on experience in use of statistical and mathematical models to analyze experimental data collected in the CRSC/Math Instructional Research Lab

2003 Affiliates Survey

Information Dissemination	Little	Moderate	Strong
Faculty Interest Database	1	4	4
Funding Alerts	1	5	3
Problem Database		6	3
Consolidated Tech Report Series	1	2	5

2003 Affiliates Survey

Event Accessibility	Little	Moderate	Strong
PPT/PDF on Web	1	1	7
PPT/PDF on CD	5	2	1
Downloadable Video	3	3	3
DVD	4	2	2
Streaming Video	4	3	1

2003 Affiliates Survey

Technology Day Topics	Little	Moderate	Strong
Statistical Software	1	4	4
Anomalies in Large Data Streams	2	2	5
Data Fusion	1	5	3
Detection of Bioterrorism	2	4	3
Theory/Methodology		4	5

2003 Affiliates Survey

Technology Day Logistics	Little	Moderate	Strong
Materials Before	2	6	1
Strong Introduction	1	4	4
Written Summary		6	3
Tutorial	3	1	5
Affiliates' Sites	1	5	2

2003 Affiliates Survey

Other Initiatives	Little	Moderate	Strong
Affiliates' Data Sets	1	3	5
Visits to Affiliates	3	3	3
Summer School	1	6	2
Shared Postdocs	3	2	4
Links to Affiliates' Web pages	1	4	4
Affiliates Problem Day		6	3

What's the Message?

- **Information Dissemination:** Interest in all items
- **Event Accessibility:** Interest only in PPT/PDF on web
- **Technology Day Topics:** Interest in all, including theory/methodology
- **Technology Day Logistics:** Particular interest in “wrappers:” strong intro, written summary [and PPT/PDF on web]
- **Other Initiatives:** Interest in all except visits, shared postdocs (but ...)

New Initiatives

- Three easy ones
 - Links from NISS web site to affiliates
 - Affiliates technical report series
 - Technology Day improvements: tutorial, intro, summary, PPT/PDF
- Three not-so-easy, but feasible, ones
 - Four specific Technology Days (knowledgeable organizers)
 - T&M Technology Days (topics; organizers)
 - Annual Affiliates Problem Day (couple with March planning meeting?)
- Three intriguing but problematic ones
 - Faculty and problem databases (hard to do well; who owns?)
 - Affiliates data sets (IP issues)
 - Shared postdocs (\$; on what terms?)