



National Institute of Statistical Sciences  
PO Box 14006, Research Triangle Park, NC 27709-4006  
Tel: 919.685.9300      FAX: 919.685.9310  
www.niss.org

## Annual Report of the Director

Alan F. Karr  
October 24, 2003

### Contents

<b>1 Highlights</b>	<b>1</b>
<b>2 NISS Research</b>	<b>1</b>
2.1 Research Initiatives . . . . .	1
2.2 Ongoing Projects . . . . .	2
2.3 Pending Proposals . . . . .	3
2.4 Planned Proposals . . . . .	3
<b>3 SAMSI</b>	<b>3</b>
<b>4 Affiliates Programs</b>	<b>4</b>
4.1 Affiliates Events . . . . .	4
4.2 Affiliate-Related Research . . . . .	4
4.3 Affiliates Services . . . . .	4
4.4 Affiliates Initiatives . . . . .	5
<b>5 Additional Activities</b>	<b>6</b>
5.1 Communication and Outreach . . . . .	6
5.2 Initiatives with Other Research Organizations . . . . .	6
<b>6 Personnel</b>	<b>6</b>
6.1 Scientific Staff . . . . .	6
6.2 Support Staff . . . . .	7
<b>7 Finances</b>	<b>7</b>
<b>8 Challenges</b>	<b>7</b>
<b>A Affiliates Events in 2003 and 2003</b>	<b>8</b>
<b>B 2003-04 Postdoctoral Fellows</b>	<b>9</b>

# 1 Highlights

NISS continues to develop along multiple dimensions. Highlights of the year include:

**Statistical and Applied Mathematical Sciences Institute (SAMSI).** An extremely successful first year at SAMSI has produced major benefits for NISS as well. These include scientific interactions—in particular, NISS is a major factor in selection and successful execution of SAMSI programs, very favorable visibility and financial stability (§3).

**Bioinformatics Research.** Under the leadership of Assistant Director Stanley Young, NISS has a growing presence in bioinformatics. Two projects (§2.1) have been initiated during the year, one in partnership with a major foundation (the Hereditary Disease Foundation (HDF) of New York) and one in collaboration with a major RTP neighbor of NISS (CIIT Centers for Health Research (CIIT)). There is also strong involvement with the Affiliates program and with SAMSI.

**Postdoctoral Program.** After a relatively down year (in terms of numbers, not quality) in 2002, the postdoctoral program blossomed in 2003, with a total of ten (an all-time record for NISS) appointments for some part of 2003–04. As important is the diversity of arrangements, which included two appointments under the Federal agency affiliates postdoctoral program, one shared with Carnegie Mellon University (CMU) as part of the digital government project (§2.2), one shared with SAMSI, one shared with the CIIT, and one funded the HDF. Another recognition of the value of the NISS postdoctoral experience is that two postdocs have accepted faculty positions, and were granted leave to come to NISS before assuming these positions in 2004.

**Affiliates Program.** Despite difficult financial times, the program is growing. A new Affiliates Proposal Development Fund (APDF) made its first two awards (§4.3). Based on a survey of the affiliates, several initiatives are underway, including an Annual Affiliates Problem Day (§4.4).

**Community Service.** Increasingly, NISS is leading initiatives of the statistical sciences community, and perhaps more important, the community is beginning to look to NISS for such leadership. Our most intensive effort has been on National Defense and Homeland Security (NDHS), which has included co-sponsoring a broad planning meeting, my chairing an ongoing committee addressing NDHS issues and a SAMSI workshop on NDHS planned for April, 2004. NISS is also a major co-sponsor of the 2004 Conference on Mathematical Methods in Reliability (MMR 2004), handling a variety of scientific, logistical and financial arrangements. NISS will host the 2008 “Interface” meeting.

## 2 NISS Research

### 2.1 Research Initiatives

**Bioinformatics.** NISS has established a vigorous research program in bioinformatics, which engages two postdoctoral fellows and interacts strongly with the SAMSI program on Data Mining and Machine Learning (DMML).

*CIIT–NISS Joint Postdoctoral Program.* NISS and CIIT have created joint postdoctoral program in bioinformatics, one-half of which is funded by each organization. (Located in Research Triangle Park, the CIIT is an independent, not-for-profit research organization dedicated to prevention of human disease. The initial thrust of the research is analysis of data from high-density microarrays. CIIT and NISS are sharing the costs (approximately \$100,000, including indirect costs) for the first year of the program, and seeking outside support for future years.

*HDF.* NISS is providing statistical and computational expertise to the HDF's Cure Huntington's Disease Initiative, a national effort that involves, in addition, biologists and chemists from a large number of institutions. Our major role is to analyze data generated by the biologists and chemists, a process that feeds back to the physical experiments. Funding for this two-year project is \$225,000.

**National Defense and Homeland Security.** Working with multiple organizations and individuals, NISS is helping to define and realize a proper role for the statistical sciences in NDHS. We co-sponsored (together with the Los Alamos National Laboratory (LANL)) a December 2002 "Planning Meeting on Statistics and National Defense and Security," at which a diverse group of 30 leaders in the statistics profession and other key individuals met to discuss the state of the statistical sciences and the field's infrastructure with respect to today's and tomorrow's national defense and security challenges. From that meeting grew an informal committee, which I chair,<sup>1</sup> that is "working" a number of issues.

A proposal for a workshop "Mathematical Sciences Research to Meet Counterterrorism Needs" was submitted in July 2003 in response to NSF's "Actions to Combat Terrorism" solicitation, but was declined because despite an explicit call for workshop proposals in the solicitation, none were funded. The thrust of the proposal was to articulate an NDHS research agenda for the statistical sciences, applied mathematics and the decision sciences, including operations research, that would yield genuine impact, in the form of six *proposal outlines* to be shared with the community in a white paper. Recognizing the need for the workshop, SAMSI has agreed to sponsor it under a new program of SAMSI Hot Topics Workshops, and it will be held at SAMSI on April 1–2, 2004. In addition, the SAMSI workshop will explore the feasibility of SAMSI's undertaking a program on NDHS in 2005–06.

## 2.2 Ongoing Projects

Over the past year, there have been three main threads in the NISS research program and one smaller effort, together with the emerging activity in bioinformatics (§2.1).

**Digital Government (DGII).** This project, *Data Confidentiality, Data Quality and Data Integration for Federal Databases: Foundations to Software Prototypes* is a four-year, \$2,560,000 follow-on to the first NISS Digital Government (DG) project. The research addresses a broad set of statistical problems with special application to unique problems faced by federal agencies—especially the necessity, in an electronic world, to balance privacy and confidentiality against user access to high-quality statistical data. Products include not only numerous conference presentations and journal articles but also software: the NISS Data Swapping Toolkit (DSTK) was released in October of 2003. Participants come from CMU, the University of Michigan, Purdue University and Southern Methodist University, as well as the five agency partners: the Bureau of Labor Statistics (BLS), Bureau of Transportation Statistics (BTS), Census Bureau (Census), National Agricultural Statistics Service (NASS) and National Center for Education Statistics (NCES).

**Computer Model Evaluation.** This effort comprises two linked projects: (1) An NSF-funded Focused Research Group (FRG) project *Statistical Framework for Evaluation of Complex Computer Models*; and (2) a targeted project funded by NISS Affiliate General Motors (GM). The research addresses Bayesian methodology for model validation.

**Information Technology,** composed of two NSF-funded projects.

The goal of the five-year project *Collaborative Research: Acquiring Accurate Dynamic Field Data Using Lightweight Instrumentation*, has its to design, implement and evaluate light-weight, in-field instru-

---

<sup>1</sup>Other members are John Bailar, Stephen Fienberg, Elizabeth Halloran, William Hanley, Sallie Keller-McNulty, Jon Kettenring, Albert Madansky, Vijay Nair and Nancy Spruill.

mentation for software testing and profiling. The project is funded by the CISE Directorate of NSF under the Information Technology Research (ITR) program. Collaborators include software engineers from the Georgia Tech, the University of Maryland, Vanderbilt University and the University of Washington.

The thrust of the project (*Bayesian Models Linking Web Site Structure and Usage*) is to relate visitor behavior to web site structure through a series of increasingly complex Bayesian models. Data are drawn from a financial services site, a content site and an information site. The principal partner is 2nd Sight, Inc., a provider of Web site analytics.

**Global Climate Change (GCC).** This project funds an Federal Agency Affiliate postdoc (§4.3) to work on statistical issues associated with global climate change. The specific foci are: 1) Trend analysis of extremes; (2) Modeling extremes in temperature; and (3) Network design for extremes.

## 2.3 Pending Proposals

Three proposals are pending as of the date of this report.

**University Countermine Research Center (UCRC).** NISS and SAMSI are both part of a Duke University proposal submitted to the Army Research Office for a “University Countermine Research Center” (UCRC) focused on developing in-depth knowledge and understanding of countermine phenomenology.

**Statistical Disclosure Limitation for Geospatial Image Data (HSD).** This proposal was submitted in July 2003 in response to the “Human and Social Dynamics” solicitation at the NSF.

**Agent-Based Simulation of Ecological Alcohol Systems (NIAAA).** This proposal has been submitted in response to the BAA “Ecosystem Models of Alcohol-Related Behavior” issued by the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Collaborators include alcohol and public policy researchers and statisticians from Buffalo State University, George Mason University (submitter of the proposal) and the University of Missouri.

## 2.4 Planned Proposals

Discussions are underway for an effort to define “the statistical computing infrastructure of the future,” with the goal of submitting a proposal in response to an anticipated cyberinfrastructure RFP from the CISE directorate at the NSF.

# 3 SAMSI

The relationship between NISS and SAMSI is now assuming a more definitive shape.

**Science.** The model of NISS as performing class-disciplinary research *driven* by scientific disciplines other than statistics (and often stimulated, informed or funded by Affiliates) and SAMSI as engaged in research in theory and methodology in the statistical and applied mathematical sciences *framed* by disciplinary science, and with each institute stimulating the other, is becoming more concrete. For example, the 2003–04 SAMSI program on DMML is taking place this year in response to strong interest among the Affiliates, is using testbed databases provided by NISS affiliates (GM and ICAGEN, and potentially others) and draws more than one-half of its participants of all kinds (workshop attendees to long-term visitors) from affiliates.

Reflecting affiliate interests as well as its own research program, NISS is a moving force behind the planned SAMSI 2004–05 program on “Latent Variable Models in the Social Sciences” and the potential 2005–06 program on NDHS (§2.1).

**Affiliates Programs.** SAMSI is a major attraction to university departments to become NISS/SAMSI University Affiliates. Because SAMSI programming reflects all affiliates' interests, SAMSI also has increasing influence on the NISS (corporate, federal agency and national laboratory) affiliates.

**Visibility.** SAMSI generates significant visibility for NISS, through workshops, research visitors and the NISS/SAMSI University Affiliates. SAMSI workshops too large to hold at NISS have at least one activity (such as a poster session) at NISS, bringing hundreds of people per year to the building.

## 4 Affiliates Programs

While still developing, the affiliates programs are assuming a distinctive form: a rich set of events (§4.1) and ongoing engagement of affiliates in NISS research (§4.2), coupled with useful services to affiliates (§4.3). Affiliates initiatives are described in §4.4.

### 4.1 Affiliates Events

Affiliates events currently consist of:

**Technology Days** that give affiliates early access to NISS research, or address important emerging areas.

**Short Courses** on important areas of theory and methodology, which have been held at both NISS and affiliates' sites.

**Workshops**, which this year will include a major workshop on risk analysis co-sponsored and hosted by Iowa State University.

**SAMSI Workshops**, many of which inform the course of NISS research.

**Co-Sponsored Events**, which in 2003 included EURANDOM workshops (especially an April 2003 workshop on data mining) and the ISI International Conference on Environmental Statistics and Health (July; Santiago de Compostela, Spain).

**Planning Meetings** twice each year, in March at NISS and in August at the JSM.

Affiliates events for 2003 and 2004 are summarized in Table 1 in Appendix A.

### 4.2 Affiliate-Related Research

During 2003, NISS was engaged in research funded by, or in collaboration with the BLS, BTS, Census, US Environmental Protection Agency (EPA), General Motors, ICAGEN, LANL, NASS and NCES.

### 4.3 Affiliates Services

**Job Listings.** A Web-based job listing service for all affiliates was initiated in October, 2001. It is available at [www.niss.org/affiliates/ajls.html](http://www.niss.org/affiliates/ajls.html).

**Publicity.** The most recent announcement about the program, listing current affiliates and inviting others to join, appeared in the December, 2002 issue of *AMSTAT News*. All affiliates received a poster-size, color version. Another announcement is planned for 2003.

**Postdoctoral Program for Federal Agency Affiliates.** A joint postdoctoral program of NISS and its Federal Agency Affiliates (FAAs) is in operation. Two appointments have been made, in conjunction with the Census Bureau and EPA. There is also strong interest from NIST.

**Event Co-Sponsorship.** Both as service to affiliates and as a means of increasing the flexibility of the Affiliates' Reimbursement Accounts, NISS is selectively co-sponsoring a number of other events. Affiliates may use their Reimbursement Accounts to attend. As part of a broader agreement with EURANDOM (§5.2), all EURANDOM workshops are part of this program. Upcoming examples include the 2004 ASA Quality and Productivity Research Conference (Raleigh; May 19–21) and MMR 2004 (Santa Fe; June 21–25).

#### 4.4 Affiliates Initiatives

**Affiliates Proposal Development Fund (APDF).** Created this year to further the NISS mission “to catalyze . . . high-impact, cross-disciplinary research involving the statistical sciences,” the APDF makes two awards annually, of up to \$10,000, for expenses associated with preparation of cross-disciplinary, preferably multi-institution proposals by researchers from NISS Affiliates or NISS/SAMSI University Affiliates. Inaugural APDF awards, selected by a committee of the Board of Trustees and announced at the NISS/SAMSI JSM Reception, went to: Kimberly Sellers (Department of Statistics, Carnegie Mellon University) and Elizabeth Arnold (Department of Mathematics, Texas A&M University) for the proposal *Getting Something from Nothing—Assessing Information Gain from Lack of Information Due to Cell Suppression*, and to Yehuda Vardi (Department of Statistics, Rutgers University), Lorraine Denby (Avaya Labs), Jean Meloche (Avaya Labs) and Donald Smith (Department of Computer Science, University of North Carolina at Chapel Hill) for the proposal *Validation of Network Tomography*.

**Affiliates Survey.** Following up on the March 2003 Affiliates Planning Meeting, an informal, web-based survey was conducted of suggestions raised at that meeting, and the results reported at the August 2003 JSM Affiliates Lunch. Notable among results was strong interest in an affiliates problem day.

**Annual Affiliates Problem Day.** The first of these will be held on March 4, 2004, in conjunction with the March 5 Affiliates Planning Meeting. At it, several corporate, federal agency and national laboratory affiliates will make in-depth presentations of research needs, with the goal of catalyzing collaborations to address them that involve NISS and other affiliates of all kinds.

**Events at Affiliates' Sites.** The affiliates survey also evidenced strong interest in affiliates events at sites other than NISS.<sup>2</sup> Although this must be done selectively, it clearly is desirable. The next such event will be the April 2004 workshop on risk analysis, which will be hosted and co-sponsored by Iowa State University.

**Retention and Recruitment.** Two new corporate affiliates in the pharmaceutical industry have joined: Aventis and ICAGEN. The only losses were one Federal agency (the BTS) and two university departments; in all three cases, where NISS' internal “champion” departed. Current membership is comprised of 13 corporations, 10 federal agencies and national laboratories, and 31 university departments.

With the assistance of the Affiliates Committee of the Board of Trustees and under the leadership of Assistant Director Thomas Gerig, a campaign has been underway since the fall of 2003 to increase the number of NISS/SAMSI University Affiliates. A parallel effort is underway to attract university mathematics departments with significant activity in applied mathematics or statistics to become NISS/SAMSI University Affiliates. Thomas Gerig and Thomas Banks (Associate Director of SAMSI) are leading this activity. A third campaign to attract new corporate affiliates, led by Assistant Director Gary McDonald, is also underway.

---

<sup>2</sup>In the past, such events have been hosted by Amgen, BLS and Telcordia Technologies.

While it seems clear that this third effort will succeed in the long run, the current difficult economic climate militates against short-run success.

## 5 Additional Activities

### 5.1 Communication and Outreach

Largely, these are expansions and refinements of ongoing activities.

**NISS Newsletter.** The quarterly newsletter is distributed by both mail and the NISS web site. Katherine Kantner continues as the editor.

**Monthly updates** from me are sent to the Board of Trustees and Members of the Corporation, with separate updates to the Affiliates.

**JSM 2003 Events.** At JSM 2003, I reported on NISS to the Board of Directors of the ASA and the IMS Council, and made a presentation on the NISS/SAMSI University Affiliates program at the ASA-organized session for academic program representatives. I also made a presentation on postdoctoral (and other) opportunities at NISS and SAMSI at the 2003 New Researchers Conference of the IMS. More than 35 affiliates attended the August JSM Affiliates Lunch, and more than 150 people came to the NISS/SAMSI Reception on Monday evening, August 4. A NISS- and SAMSI-sponsored session “SAMSI After One Year” attracted an audience of approximately 200.

**Sacks Award.** The 2003 Jerome Sacks Award for Cross-Disciplinary Research was presented to Raymond Carroll of the Department of Statistics at Texas A&M University. At current low interest rates, the \$30,000+ Sacks Award endowment is not able to support the award of \$1000; the difference is made up from NISS general funds.

### 5.2 Initiatives with Other Research Organizations

**EURANDOM.** A Memorandum of Agreement with EURANDOM has been implemented to establish and implement sensible forms of interaction, including joint workshops and a postdoctoral exchange program, both of which are underway already.

## 6 Personnel

Flexible (medical and dependent care) spending accounts have been instituted as an additional personnel benefit.

### 6.1 Scientific Staff

The NISS scientific staff is strong and stable. It includes:

**Thomas Gerig**, Assistant Director for NISS/SAMSI University Affiliates, has as primary responsibility communication with and recruitment of NISS/SAMSI University Affiliates.

**Gary McDonald**, Assistant Director for Program Development, is playing a major role in expanding NISS research on computer model evaluation (including possible creation of a long-term capability), especially in connection with affiliates. He plays a significant role in the SAMSI 2003–04 program on DMML.

**Ashish Sanil** is Research Statistician at NISS. He has a major role in research on DG, data quality (DQ) and software engineering, and is co-PI on the WEB project (§2.2).

**Stanley Young**, as Assistant Director for Bioinformatics, is managing and developing research programs in drug discovery, genomics and proteomics, as well as interaction with affiliates in the pharmaceutical industry. Reflecting affiliate interests, he is also a working group leader in the SAMSI DMML program.

**Postdoctoral Fellows.** As noted in §1, the postdoctoral program increased dramatically in terms of both number and diversity of appointments. The complete set of postdocs for 2003–04 appears in Appendix B.

## 6.2 Support Staff

The NISS staff consists of three extremely dedicated and efficient individuals. **Martha Williamson**, Administrative Assistant, provides strong support for me and functions as business manager of NISS. **Katherine Kantner** has major responsibilities in connection with the Affiliates Program, the Board of Trustees and communication for both NISS and SAMSI. **James Thomas**, now full-time computer system manager, provides system, software and network support to NISS and SAMSI.

## 7 Finances

NISS was financially stable in FY 2003 (ended June 30, 2003) and will be so again in FY 2004.

## 8 Challenges

Not surprisingly, the 2003 NISS Annual Meeting will focus on the biggest challenge facing NISS: development and then implementation of a strategic vision that:

- Creates a clear and valued identity for NISS within the statistical sciences community;
- Promotes broader and larger-scale engagement of the statistical sciences community (as various scientific communities) in NISS' cross-disciplinary research program, increases the scale of the program and maintains its quality;
- Continues and increases the value of the affiliates programs to both the affiliates and NISS;
- Maintains NISS' leadership in career development, especially for postdoctorals;
- Creates effective, stable, definable relationships between NISS and other organizations, including SAMSI, affiliates, our parent societies and other national organizations such as CATS and CNSTAT.

## A Affiliates Events in 2003 and 2004

Event	Date	Location	Affiliate Attendance
<u>2003</u>			
Technology Day: Statistical Issues in Proteomics	3/6/03	RTP	50
Affiliates Planning Meeting	3/7/03	NISS	30
Technology Day: Internet Tomography	3/28/03	NISS	25
Computer Experiments Short Course	6/13/03	GM	50
JSM Affiliates Meeting	8/3/03	San Francisco	40
SAMSI DMML Kickoff Workshop	9/7–9/03	RTP	40
SAMSI Internet Tomography Workshop	10/13–15/03	RTP	30
SAMSI Congestion and Heavy Traffic Modeling Workshop	10/31-11/1/03	RTP	
Computer Experiments Short Course	11/7/03	GM	50
SAMSI Undergraduate Workshop	11/14–15/03	NISS	30
Data Mining—Handling the Flood of Data			
Data Confidentiality Technology Day	11/03	Washington	20
<u>2004</u>			
SAMSI MULTI Kickoff Workshop	1/19–21/04	RTP	
SAMSI Workshop on Multi-Scale Phenomena in Soft-Matter, Nano-Materials	2/15-017/04	RTP	
Affiliates Problem Day	3/4/04	NISS	
Affiliates Planning Meeting	3/5/04	NISS	
SAMSI Workshop on NDHS	4/1–2/04	RTP	
SAMSI Workshop on Fluctuations and Continuum Equations in Granular Flow	4/15–17/04	RTP	
Workshop on Overarching Issues in Risk Assessment	4/22–24/04	Iowa State University	
QPRC 2004	5/19-21/05	Raleigh	
SAMSI DMML Closing Workshop	5/20–22/04	RTP	
MMR 2004	6/21-25/04	Santa Fe	
Technology Day: Bayesian Strategies for Computer Model Evaluation	Spring 2004	NISS	

Table 1: Summary of principal affiliates events in 2003 and 2004.

## **B 2003–04 Postdoctoral Fellows**

**John Aston** (Ph.D., Statistics, University College London), now in his second year at NISS, is based full-time at and supported by the Census Bureau, where he is working in the Time Series Research Group of the Statistical Research Division.

**Max Buot** (Ph.D., Statistics, Virginia) will spend January–May of 2004 at NISS. Max has accepted a postdoctoral position at Carnegie Mellon University, working with Stephen Fienberg on digital government, and will visit NISS while Steve is on sabbatical in Paris.

**Jun Feng** (Ph.D., Computational Chemistry, North Carolina) is supported by the Hereditary Disease Foundation (HDF) (§2.1).

**Shanti Gomatam** (Ph.D., Statistics, Florida State), who has left NISS to accept a position of the FDA, led NISS research on data swapping and participated on work on regression servers. Her research was supported by the NISS digital government project and the NCES.

**Amy Grady** (Ph.D., Statistics, North Carolina) is doing research pertaining to global climate change, focusing on trend analysis of extremes of precipitation and temperature, modeling extremes in temperature and network design for extremes. She is supported by a cooperative agreement with the EPA.

**Murali Haran** (Ph.D., Statistics, Minnesota) arrives at NISS in September for a year, after which he will assume a faculty position in statistics at Penn State. He is working on the software engineering and Web data projects (§2.2), both of which are funded by the NSF.

**Xiaodong Lin** (Ph.D., Statistics, Purdue) divides his time between the SAMSI DMML program and NISS, where he will be involved with digital government, working on privacy-preserving data integration. Xiaodong will assume a full-time faculty position at the University of Cincinnati in 2004.

**Kejun (Jack) Liu** (Ph.D., Bioinformatics, North Carolina State) is the first CIIT–NISS Joint Postdoctoral Fellow. He is mentored by Stan Young and Kevin Gaido of CIIT, and divides his time between CIIT and NISS. He is also involved in the SAMSI DMML program.

**Rui Paulo** (Ph.D. Statistics, Duke) has shifted from a split NISS/SAMSI appointment in 2002–03 to a full-time appointment at NISS for 2003–04, working on GM- and NSF-supported research on computer model evaluation.

**Danny Walsh** (Ph.D., Statistics, Washington) has also moved from SAMSI to NISS, and is involved in GM- and NSF-funded research on computer model evaluation.