

## NISS Holds 2005 Annual Meeting

The 2005 NISS Annual Meeting was held on Friday and Saturday, November 4 and 5, at NISS headquarters. Nearly all members of the Board of Trustees and NISS Corporation were able to attend. Vijay Nair, chair of the Board of Trustees, ran the meeting with his usual alacrity.

A highlight of the meeting was a review of the NISS research program. This featured overviews by director Alan Karr, associate director Nell Sedransk—emphasizing on NISS-SAMSI relationships and assistant director Stanley Young, focusing on bioinformatics. The substantive portion consisted of presentations of their research by the four current NISS postdocs:

- Michael Last: Light-Weight Instrumentation of Fielded Software
- Anna Oganian: SDC Methods and their Combinations
- Francisco Vera: Secure Statistics Software
- Mi-Ja Woo: New Measures of Data Utility

Another important activity was discussions of activities and plans of the principal committees of the Board of Trustees:

- The Affiliates Committee, chaired by Lee Wilkinson, who reported on very successful "targetted" efforts to recruit new affiliates.
- The Awards Committee, chaired by Jessica Utts, whose proposal to establish NISS New Researcher Fellowships, was approved.
- The Engagement and Outreach Committee, chaired by Sally Morton, which is working with Karr and Sedransk to make 2006 the year of engagement and outreach for NISS.
- The Science Committee, chaired by John Rolph, which is continuing to seek new scientific opportunities for NISS.



NISS Board Chair,  
Vijay Nair

A final highlight was consideration of plans for a possible expansion of the NISS building to accommodate both a larger NISS and a larger SAMSI, which included presentations by Karr and Kevin Montgomery of O'Brien Atkins (OBA), the architects for the existing, award-winning building. Following the annual meeting, the building committee, chaired by Jon Kettenring, met by teleconference, and recommended to the Executive Committee that NISS engage OBA for design of the expansion. In December of 2005, the Executive Committee approved this recommendation, and the process is moving forward.

### On the Inside

Director's Corner: Page 2  
Postdoc Announcement and  
Turkey Fry Photos : Page 3  
Research News: Page 4  
NISS and SAMSI Calendars: Page 5  
Affiliated Organizations: Page 6

## Director's Corner

Alan F. Karr



### The NISS/SAMSI Partnership

The initial concrete steps to create SAMSI took place in July of 2000, the same month I became director of NISS. Innumerable meetings, pizzas and late nights later, and five minutes before the deadline, the proposal was submitted. The rest, to use a trite phrase, is history. But of course it is living history that energizes me constantly.

At the November 2000 NISS annual meeting, there was some skepticism about whether SAMSI would be good for NISS. Or would it somehow dilute the impact of our research or distract us from our core missions of cross-disciplinary research, postdoctoral and the then-nascent affiliates program?

Nearly a year ago, SAMSI went through a very successful third-year review, the most important consequence of which is being encouraged to submit a renewal proposal to fund SAMSI for 2007-12, and at a level significantly higher than the current \$2 million per year. So this is a reasonable time to ask whether SAMSI has been as good for NISS as I had hoped in 2000, and to think about how the partnership should evolve.

The short answer is that in every way the NISS-SAMSI partnership is working, and in most cases better than I had dared dream. And despite some daunting implications about expanding the building, as I said at this year's annual meeting, it is unthinkable that NISS and SAMSI should not remain and thrive together.

That SAMSI would contribute to the financial stability of NISS was always clear. This matters, but it alone is not a very persuasive reason for NISS to invest so much in SAMSI.

The identity of NISS is such a reason, however, and the most important aspect of the partnership is that it has let NISS concentrate on being NISS (in other words, on our core missions), rather than trying to be both NISS and SAMSI.

Our research on data confidentiality (see page 4) and our deep and growing involvement in ESSI-Stat and NESSI (the new Education Statistics Services Institute-Statistics and NAEP Education Statistics Services Institute) epitomize NISS' role in working important problems at both cross-disciplinary and industry/government-academia interfaces. These are not roles that SAMSI would or should undertake, but the more important thing is that SAMSI frees NISS to do them exceedingly well.

This is not to say that NISS science and SAMSI science do not intersect. Indeed, the vision since the beginning is that NISS would be both a generator and "follow-upper" of SAMSI programs. The first of these, in particular through the affiliates programs, has worked stunningly well. SAMSI programs on Data Mining and Machine Learning (2003-04), Internet Traffic (2003-04), Latent Variable Methods in the Social Sciences (2004-05), National Defense and Homeland Security (2005-06), Computer Models (2006-07) and risk analysis (planned for 2007-08) respond directly to affiliate interest, so that SAMSI enables NISS to serve our affiliates.

The follow-up role is also working, albeit more slowly. An ongoing NISS project on social networks grows from the SAMSI LVSS and NDHS programs. Others are in the works.

SAMSI has also been a boon to postdoctoral recruiting, making for a richer, more varied experience for NISS (and SAMSI!) postdocs. The research note on page 4 describes work catalyzed in large part by a shared postdoc.

And of course SAMSI has been a major attraction in building and maintaining the affiliates programs, especially the NISS-SAMSI University Affiliates Program.

Finally, SAMSI has increased the visibility of NISS dramatically. More than a thousand researchers per year come to NISS/SAMSI. And it is impossible to overestimate how good it feels to be in a building full to the point of overflowing with such outstanding statistical scientists!

But the "overflowing" raises sobering questions for NISS. Both institutes need more—and more effective—space, and the Board of Trustees has approved expenditure of funds on the initial design phase for an expansion of our building, as recommended by Nell Sedransk and me. Although SAMSI will continue to pay rent, if the project comes to reality, it will entail sizable financial commitments by NISS as well.

As I said before, it is unthinkable not to make these commitments. There is too much at stake, for NISS, for SAMSI, and for the profession, for us not to move ahead with enthusiasm and optimism. I hope for your support, and I welcome your ideas about how to make this unique partnership even better.

#### NISS Newsletter

Katherine Kantner, Editor

Published quarterly by the

National Institute of Statistical Sciences

PO Box 14006

Research Triangle Park, NC 27709

919.685.9300

[www.niss.org](http://www.niss.org)

© 2006, National Institute of Statistical Sciences

All Rights Reserved

# Postdoctoral Fellowships

NISS is accepting applications from statistical, computer and applied mathematical scientists for postdoctoral positions. Several appointments are anticipated, with starting dates in summer 2006 or sooner. Appointments will be for two years, with extensions possible. The salary is currently \$65,000 per year.

In accordance with the NISS mission, postdoctoral fellows participate in cross-disciplinary research projects in collaboration with statisticians and disciplinary scientists from universities, industry, national laboratories and government agencies. They are mentored by senior statisticians and scientists, and interact strongly with one another, as well as with researchers at NISS' sister institute, the Statistical and Applied Mathematical Sciences Institute (SAMSI).

Postdoctorals will engage in NISS projects in the areas of *digital government*, especially the intersections among data confidentiality, data integration and data quality; *dynamics of social networks*, *information technology*, including software engineering, *education statistics*—NISS is a primary partner of the American Institutes for Research (AIR) in operating the Education Statistics Services Institute—Statistics (ESSI-Stat), *bioinformatics*, including the integration of genomics, proteomics and metabolomics, and *data quality*. There is also time available for independent research.

Other opportunities arise from the NISS Affiliates Program. NISS and its Federal agency affiliates (the Bureau of Labor Statistics, Census Bureau, National Agricultural Statistics Service, National Cancer Institute, National Center for Education Statistics, National Center for Health Statistics, National Institute of Standards and Technology, and National Security Agency) operate a joint postdoctoral program under which NISS appoints postdoctoral fellows for 2-3 year terms in challenging, exciting assignments, principally at agency locations in Washington, DC.

Shared appointments are also possible with SAMSI; see [www.samsi.info](http://www.samsi.info) for information about SAMSI programs planned for 2006-07.

Criteria for selection include demonstrated research ability in statistics or a relevant scientific discipline, interest and—to a lesser degree, experience—in project areas, strength in computation, skill in verbal and written communication, and genuine commitment to cross-disciplinary research.

Applicants must have received, or expect to complete, a doctorate in 2000 or later. Women and members of minority groups are particularly encouraged to apply.

Applications should consist of a letter of interest containing addresses, telephone numbers and citizenship status; three letters of reference; a one-page biographical sketch; a publication record or dissertation abstract; and a one-page statement of interest responding to the criteria and projects listed above. These items are to be submitted by E-mail, as PDF files if possible, to [postdoc-application@niss.org](mailto:postdoc-application@niss.org).

The deadline for full consideration is January 31, 2006. Later applications will be considered as resources permit. Appointments may be made at any time.

*NISS is an AA/EEO employer*

## Good Eats at the Annual Turkey Extravaganza !



# Sharing Information without Sharing Data

## Research Note

Increasingly, databases containing related data reside in multiple locations, under the control of multiple owners. These owners are often unwilling or unable to share the data with one another. For example, corporate data may be proprietary, and data held by government agencies may have been collected under a promise of confidentiality.

There may, however, be clear benefits to science or society from performing statistical analyses on the “integrated data.” In the example below, we show how pharmaceutical companies can benefit from an analysis that, in effect but not in reality, integrates their chemical libraries. Early identification of adverse drug effects, disease outbreaks and potential terrorist attacks illustrate social benefits. Despite federal strictures, sharing of scientific data is not routine. And sharing of much state- and district-level education data, even though it would permit more powerful and informative analyses, is rare-to-non-existent.

For the past two years, NISS has been developing methodology and software tools that address this problem, and which allow valid statistical analyses to be conducted on distributed databases without ever integrating those databases, and without use of trusted third parties (human or electronic).

This research was catalyzed by Xiaodong Lin, a joint NISS-SAMSI postdoc now at the University of Cincinnati. Other principal participants have been Ashish Sanil (former NISS research statistician, now at Bristol-Myers Squibb), Jerome Reiter (Duke), NISS postdocs Jun Feng and Francisco Vera, graduate intern William “Jimmy” Fulp (Carnegie Mellon) and NISS assistant director Stanley Young. A series of papers has resulted, all of which are on the NISS web site, at [www.niss.org/dgii/techreports.html](http://www.niss.org/dgii/techreports.html).

The techniques are based on concepts of secure multi-party computation from the computer science literature. “Secure summation” illustrates: suppose the databases owners simply wish to add values, so that each knows the sum but not the others’ values. To do so, one owner generates a large random number, adds its value,

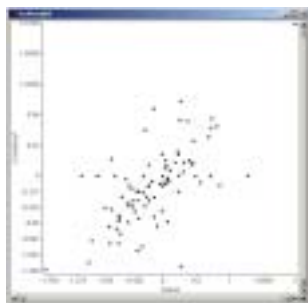
passes the sum to the second, which adds its value and passes the new sum to the third, and so on. When the sum returns to the first owner, it subtracts the random number and shares the result.

Because secure regression for horizontally partitioned data can be reduced to secure summation (for the  $X$ -transpose- $X$  and  $X$ -transpose- $Y$  matrices), we can do secure regression. (For horizontally partitioned data, data subjects rather than attributes are distributed among the owners.) Here is an example, taken from the paper “Secure Analysis of Distributed Chemical Databases without Data Integration,” by A. F. Karr, J. Feng, X. Lin, J. P. Reiter, A. P. Sanil, and S. S. Young, which appeared in the November 2005 issue of the *Journal of Computer-Aided Molecular Design*.

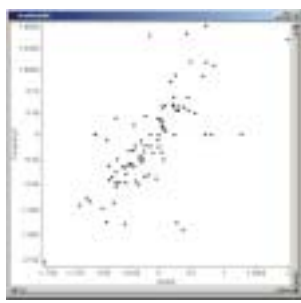
Four pharmaceutical companies have related libraries of chemical molecules, in which the response is water solubility and the predictors are a constant and 90 molecular features. The numbers of molecules in the libraries are 499, 572, 16 (!), 231, and for each company, there are features not present in its library, so that the overall analysis can only be performed by the companies cooperating. The first three scatterplots show how each of company 1, 2 and 4’s own coefficients (y-axis) related to the global (all four companies) regression coefficients (computed using secure regression, and on the x-axis). Clearly each company learns a great deal!

Since company 3, with only 16 molecules cannot even do the regression on its own, it is natural to ask whether companies 1, 2 and 4 benefit from allowing it to participate. The final scatterplot shows the {1,2,4}-regression coefficients (y-axis) plotted against the global coefficients. There are differences, because company 3 has features the others don’t.

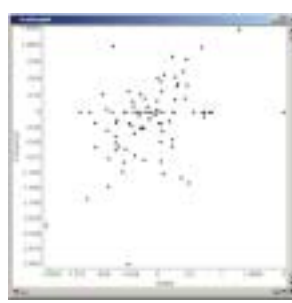
The initial version the NISS Secure Statistical Computation System (SSCS) will be released soon. Watch the web site for details!



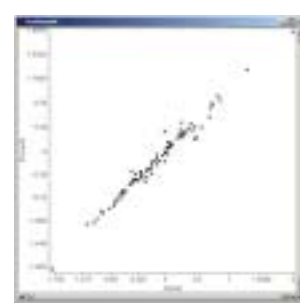
Company 1



Company 2



Company 4



Except 3

# NISS and SAMSI Calendars

## **Affiliates Planning Meeting**

March 2006 (Location and Date TBA)

## **Workshop on Survey Costs**

April 18-19, 2006 (Washington, DC)

## **Workshop on Statistical Disclosure Limitation for Geospatial Data**

Spring 2006 (NISS)

## **SAMSI Events**

### **Financial Mathematics, Statistics & Econometrics**

#### **Model Uncertainty Workshop**

January 27, 2006 (SAMSI)

### **National Defense and Homeland Security**

#### **Anomaly Detection Mid-Program Workshop**

February 3, 2006 (NCHS, Beltsville, MD)

### **Financial Mathematics, Statistics & Econometrics**

#### **Transition Workshop**

February 27-28, 2006 (Radisson RTP)

### **National Defense and Homeland Security**

#### **Social Networks Mid-Program Workshop**

March 3, 2006 (Carnegie Mellon University, Pittsburgh)

### **Two-Day Undergraduate Workshop**

March 3-4, 2006 (SAMSI)

### **National Defense and Homeland Security**

#### **Transition Workshop**

May 15-16, 2006 (SAMSI)

### **Interdisciplinary Workshop for Undergraduates**

May 22-26, 2006 (SAMSI)

### **Astrostatistics**

#### **Transition Workshop and SCMA VI**

June 12-15, 2006 (Penn State)

### **2006 Summer Program on Multiplicity and Reproducibility in Scientific Studies**

July 10-28, 2006 (SAMSI)

### **Astrostatistics**

#### **Transition Workshop and PHYSTAT**

July 15-20, 2006 (BIRS)

## **NISS Affiliates**

### **Corporations**

Avaya Labs (Basking Ridge, NJ)  
Aventis Pharmaceuticals (Bridgewater, NJ)  
Bell Labs - Lucent Technologies (Murray Hill, NJ)  
General Motors (Detroit, MI)  
GlaxoSmithKline (RTP, NC and Collegeville, PA)  
ICAGEN, Inc. (Durham, NC)  
Merck & Co. (West Point, PA)  
Metabolon, Inc. (Research Triangle Park, NC)

MetaMetrics (Durham, NC)  
Nuevolution (Copenhagen, Denmark)  
RTI International (Research Triangle Park, NC)  
SAS Institute (Cary, NC)  
SPSS, Inc. (Chicago, IL)  
Telcordia Technologies (Piscataway, NJ)  
Wyeth Pharmaceuticals (Collegeville, PA)

### **Government Agencies and National Laboratories**

Bureau of the Census (Washington, DC)  
Bureau of Labor Statistics (Washington, DC)  
Bureau of Transportation Statistics (Washington, DC)  
Los Alamos National Laboratory (NM)  
National Agricultural Statistics Service (Fairfax, VA)

National Center for Education Statistics (Washington, DC)  
National Center for Health Statistics (Hyattsville, MD)  
National Institute of Standards and Technology (Gaithersburg, MD)  
National Security Agency (Fort George Meade, MD)

## **NISS/SAMSI University Affiliates**

University of California - Berkeley (Statistics)  
Carnegie Mellon University (Statistics)  
Duke University (Statistics and Decision Sciences, and Mathematics)  
Emory University (Biostatistics)  
University of Florida (Statistics)  
Florida State University (Statistics)  
George Mason University (Statistics)  
University of Georgia (Statistics)  
University of Illinois Urbana-Champaign (Statistics)  
Iowa State University (Statistics)  
University of Iowa (Statistics)  
Johns Hopkins University (Mathematical Sciences)  
University of Michigan (Statistics and Biostatistics)  
University of Missouri - Columbia (Statistics)  
North Carolina State University (Statistics and Mathematics)  
University of North Carolina at Chapel Hill (Biostatistics, Stat-OR, and Mathematics)  
Oakland University (Mathematics and Statistics)  
Ohio State University (Statistics)  
Pennsylvania State University (Statistics)  
Purdue University (Statistics)  
Rice University (Statistics)  
Rutgers University (Statistics)  
University of South Carolina (Statistics)  
Southern Methodist University (Statistical Science)  
Stanford University (Statistics)  
Texas A&M University (Statistics)  
Virginia Commonwealth University (Biostatistics)