Final Program Report: National Defense and Homeland Security

1. Objectives

The SAMSI program on National Defense and Homeland Security achieved its principal purposes of identifying research promising paths for the statistical sciences, applied mathematics and decision sciences in problems of National Defense and Homeland Security (NDHS), and initiating research on them. This effort was especially important because previous efforts by these communities had failed to create a self-sustaining research momentum on NDHS. In addition, there have been few research efforts that had spanned the statistical sciences, the applied mathematical sciences and the decision sciences.

2. Working Groups

Four Working Groups operated throughout the year, whose principal function, as in all SAMSI programs, was to organize the research and ensure communication:

- **Agricultural Systems,** led by Mette Olufsen, Faculty Fellow (Applied Mathematics, NCSU). Barrett Slenning (College of Veterinary Medicine, NCSU) is providing significant leadership as well.

- **Anomaly Detection,** led by David Dickey, Faculty Fellow (Statistics, NCSU), and Douglas Kelly, Faculty Fellow (Statistics, UNC).

- **Data Confidentiality,** led by Lawrence Cox (NCHS) and Alan Karr (NISS).

- **Social Networks,** led by David Banks (Statistics, Duke).

Each working group had significant external participation. Specifically,

- Greg Rempala and Ryan Gill (University of Louisville) were participants in the **Agricultural Systems** Working Group, and Michelle Lacey (Tulane University), who was at SAMSI for the fall of 2005, was a regular participant.

- Deepak Agarwal (AT&T Research), Howard S. Burkom (Johns Hopkins University Applied Physics Laboratory), Kevin Ward Drummey (National Security Agency (NSA)), Joe Fred Gonzalez, Jr. (National Center for Health Statistics (NCHS)), Myron Katzoff (NCHS), James Lynch (University of South Carolina), Ted Norminton (Carleton University), Cheolwoo Park (University of Georgia), Henry Rolka (Centers for Disease Control and Prevention (CDC)), and Galit Shmueli (University of Maryland) were regular participants in the **Anomaly Detection** Working Group.

- Lawrence Cox (NCHS), Joe Fred Gonzalez (NCHS) and James Lynch (University of South Carolina, who spent the spring semester at SAMSI) were regular participants in the **Data Confidentiality** Working Group.

- Deepak Agarwal (AT&T Research), Edoardo Airoli (Carnegie Mellon University), Hugh Chipman (Arcadia University), Stephen Fienberg (Carnegie Mellon University), Myron Katzoff of the NCHS, and Ted Norminton (Carleton University), were regular participants in the **Social Networks** Working Group.

3. Outcomes

3.1 Program Level

Outcomes at the program level were:

- Significant research accomplishments by the working groups, leading to papers submitted during or after the program year.

- Formation of new collaborations, leading to proposals and research in following years. See §6.2 for an example.

- Extremely positive career impact on participants, especially postdoctoral researchers.

- Strong community interest in the program, leading to engagement in the form of research visits and workshop participation.

3.2 Working Groups

Each working group produced a "sound bite" summary of its plans and identified one or more outcomes that it would consider to be major successes. Not all did so, but each addressed high-level goals. Each working group also developed a detailed research agenda that both focuses energy and defines measures of success.

Agricultural Systems. The group set as a major goal and success identification of state of the art model predicting spread of Foot and Mouth Disease (FMD) that can be applied to study of the disease in North Carolina and the US.
Participants from SAMSI and the Research Triangle were Amit Apte (postdoc, UNC), Ping Bai (student, UNC), David Banks (Duke), Thomas Banks (NCSU), Ricardo Cortez (Tulane University; currently housed at UNC), Sava Dediu (postdoc, SAMSI), Anjela Govan (student, NCSU), Christopher Jones (SAMSI and UNC), Michelle Lacey (Tulane University; currently housed at Duke), Michael Last (postdoc, NISS), Hoan Nguyen (postdoc, NCSU), Mette Olufsen (NCSU), Barrett Slenning (NCSU) and Ralph Smith (NCSU). Thrusts of the research included:
Background presentation "Agricultural Disasters, Natural or Not: Risks and Readiness" by Slenning. Literature search for mathematical and statistical models for FMD. Review and discussion of items identified by the literature search. An introduction to Jackson networks.
As described in §6.1, this group remained active through March of 2007, producing a major multiple-author paper.

Anomaly Detection. A central problem in NDHS is the early detection of events. These might include early signs of a disease outbreak (natural or terrorist induced), imminent disastrous weather events, or any form of terrorist attack. As the nation thinks about strategies for dealing with these issues, the Working Group is trying to surveying statistical methods that might be part of a plan to deal with some of these threats.
The group created a web site with descriptions of statistical methods applicable to the kinds of data involved in NDHS issues. Examples would include time series, issues of type I versus type II error tradeoffs, outlier and change point detection.
SAMSI-and Research Triangle-based participants were David Banks (Duke), M. J. Bayarri (SAMSI), Gauri Datta (University of Georgia, visiting SAMSI), Lisa R. Denogean (postdoc, SAMSI), David A. Dickey (NCSU), Joyee Ghosh (student, Duke), Shenek Heyward (student, NCSU), Yajun Mei (Fred Hutchinson

Cancer Center, visiting SAMSI), Bahjat Qaqish (UNC), Curtis Storlie (postdoc, NCSU/SAMSI) and Francisco Vera (postdoc, NISS and SAMSI).

Research thrusts:

- Study of syndromic surveillance, which is roughly described as statistical analysis of indicators from many sources, for example, admissions to several hospitals for flu-like symptoms as well as sales of over-the-counter products related to the flu. One particular technique, the SCAN statistic, was the focus of some effort and some thinking about how to modify it to take into account background effects.

- Review of related statistical topics, such as time series, outlier detection, sequential testing methods, and multiple testing in a Bayesian framework.

- Cross-fertilization of ideas has been very important. To illustrate, Dickey presented a time series of American Airlines stock volume as part of a time series presentation, and Lynch immediately did some additional analyses that were presented to the group. An initially unsuccessful proposal produced by members of this group is discussed in §6.2.

Data Confidentiality. Because of increasing availability of external databases, construction of public release databases by federal agencies is verging on the impossible. A long-term goal is to produce an integrated suite of software tools that customize disclosure limitation strategies to statistical characteristics of databases, prolonging by years the ability of the agencies to release useful microdata.

SAMSI-and Research Triangle-based participants were Ping Bai (student, UNC), Lisa R. Denogean (postdoc, SAMSI), Joyee Ghosh (student, Duke), Alan Karr (NISS), Robin Mitra (student, Duke), Anna Oganian (postdoc, NISS), Bahjat Qaqish (UNC), Jerome Reiter (Duke), Saki Kinney (student, Duke), Francisco Vera (postdoc, NISS and SAMSI) and Mi-Ja Woo (postdoc, NISS).

Visitors during the spring semester were James Lynch (South Carolina) and Xiaodong Lin (Cincinnati; supported by NISS).

Thrusts of the research:

- Development of techniques that combine two or more methods for statistical disclosure limitation (SDL), resulting in disclosure risk and data utility performance superior to either method alone.

- Construction of new utility measures for numerical microdata, based on propensity scores and clustering.

- Exploration of SDL for data satisfying constraints such as positivity. Utility measures that account for analyses of released microdata that accommodate whatever SDL measures were applied.

- Methods and software systems for statistical analysis of distributed databases with actual integration of the data.

- A new formulation of data swapping that randomizes which attributes are swapped as well as which records are selected for swapping.

- As described in §6.1, this working group remains fully active, now as a component of the NISS affiliates program.

Social Networks. Social network statisticians have developed a rich class of static models. But new challenges that have arisen in counterterrorism, communications networks, and disaster response show the need for robust, reliable models that capture network dynamics. A major success would development of a body of theory for dynamic network models and the benchmarking of such models against applications from many different fields.

In addition to external participants listed in §2, participants based at SAMSI or in the Research Triangle are David Banks (Duke), Thomas Banks (NCSU), Sava Dediu (postdoc, SAMSI), Chung-Chien Hong (student, NCSU), Alan Karr (NISS), Michael Last (postdoc, NISS), Negash Medhin (NCSU), Hoan Nguyen (postdoc, NCSU) and Eric Vance (student, Duke).

Activities of the group:

- Development of multiple models that introduce dynamics into social networks.

- Organization of and participation in a Board on Mathematical Sciences and their Applications (BMSA) Workshop on Statistics in Networks (Washington, September 26–27) and DARPA Workshop on Virtual Worlds and Social Networks (Washington, October 18–19).

- Review of latent space models and ideas for generalization to dynamic network models, of optimization methods for dynamic networks, of estimation of homophily issues and of dynamic process models.

- Application of network models to Hurricane Katrina evacuee surveys.

- Preparation of a research proposal on network models, collaboratively with AT&T Labs, which will allow models to be tested on real data.

- Planning and preparation of papers for a special issue of *Computational Mathematics and Organizational Theory*.

Main research directions:

- Development of a framework for characterizing behavior of dynamic social network models, leading to a paper that has been accepted for publication (Banks et al., 2005).

- Optimization models for social networks (Hong and Medhin, 2006 a,b).

- Testing multiple dynamic network models on data from Enron, AT&T and Hurricane Katrina.

- Assessing agent-based models vis-à-vis dynamic versions of p models (Vance et al., 2006).

4. Workshops

4.1 Kickoff Workshop

The September 11–14, 2005 kickoff workshop for the program attracted more than 100 attendees, and met the stated goal of informing the composition and activities of the Working Groups. Details of the program are at www.samsi.info/workshops/2005ndhs-workshop200509.shtml.

4.2 Mid-Program Workshops

Three of the working groups held off-site mid-program workshops, which were extremely successful in attracting attendees who are not working group participants, as well as bringing together in-Triangle and out-of-Triangle participants. Specifics are as follows:

Anomaly Detection. The workshop was held on February 3, 2006 at the National Center for Health Statistics in Hyattsville, MD. Attendance was approximately thirty. Keynote presentations were made by Donald E. Brown (Systems and Information Engineering, University of Virginia), Howard S. Burkom (National Security Technology Department, Johns Hopkins Applied Physics Laboratory) and Carey E. Priebe (Applied Mathematics and Statistics, Johns Hopkins University). Working group participants making presentations were Gauri Datta, David Dickey, Ryan Gill, Michael Last and Francisco Vera.
Full details are available at www.samsi.info/200506/ndhs/workinggroup/ad/Workshop.htm.

Data Confidentiality. This workshop was held on March 13, 2006, also at the National Center for Health Statistics in Hyattsville, MD, and attracted more than 45 attendees, including researchers from the Bureau of Labor Statistics, Census Bureau, Energy Information Administration, National Center for Education Statistics and National Center for Health Statistics. Principal presentations were by Lawrence Cox (NCHS), Jay Kim (NCHS) and Aleksandra Slavkovic (Penn State). Presentations on research conducted by the working group were made by Lisa Denogean (SAMSI), Anna Oganian (NISS), Mi-Ja Woo (NISS) and Francisco Vera (NISS and SAMSI). James Lynch and Jerome Reiter led discussion sessions.
Full details are available at www.samsi.info/200506/ndhs/workinggroup/dc/workshops/midyear workshop.htm.

Social Networks. The workshop was held on March 2, 2006 at Carnegie Mellon University, the principal location of non-Triangle participants in the working group. Attendance was approximately 25, including Canadian participants Shirley Mills and Ted Norminton. Presentations were made by Alan Karr (NISS), Purnamrita Sarkar (CMU), Anna Goldenberg (CMU), Eric Xing (CMU), Eric Kolaczyk (Boston University), Steve Henneke (CMU) and Eric Vance (Duke University).

4.3 Transition Workshop

An informal transition Workshop for the NDHS program was held in Research Triangle Park, NC, in October, 2006, in conjunction with the Army Conference on Applied Statistics (ACAS), which was be hosted by NISS. Three sessions at the ACAS contained presentations resulting from the NDHS program, informing the nearly 100 attendees about the achievements of the program.

5. Education and Outreach

5.1 Education and Outreach

A seminar course on National Defense and Homeland Security took place in the fall of 2006 at SAMSI, led by Alan Karr. Eight students were enrolled.

5.2 Undergraduate Workshop

An undergraduate workshop based on the National Defense and Homeland Security program took place on March 3–4, 2006. There were approximately 30 attendees, one-half of them women. Presentations were made by David Banks, Negash Medhin, Hoan Hguyen, Eric Vance, Ping Bai, Sava Dediu, Anjela Govan, Lisa

Denogean, Francisco Vera, David Dickey, and Alan Karr. Details are available at
www.samsi.info/workshops/2005ug-workshop200603.shtml.

6 Follow-On Activities

6.1 Continuing Research

The **Agricultural Systems** Working Group remained active through March, 2007, producing a paper entitled
"Stochastic and Deterministic Models for Agricultural Production Networks" by P. Bai, H. T. Banks, S. Dediu,
A. Y. Govan, M. Last, A. Lloyd, H. K. Nguyen, M. S. Olufsen, G. Rempala, and B. D. Slenning,[1] which has
been submitted to *Mathematical Biosciences and Engineering* (Bai et al., 2007).

The **Data Confidentiality** Working Group became a working group of the NISS Affiliates Program,
and continues to meet weekly. It is led by Alan Karr, directorate liaison to the NDHS program. Program
participants active in it include L. Cox (NCHS), J. F. Gonzalez (NCHS), A. Karr (NISS), M. Katzoff (NCHS),
J. Reiter (Duke) and F. Vera (Clemson). Multiple publications catalyzed by the SAMSI working group were
completed Denogean et al. (2006); Ghosh et al. (2006); Karr et al. (2006b,a); Oganian and Karr (2006); Sanil
et al. (2004); Woo et al. (2006).

6.2 FRG Proposal

Several members of he **Anomaly Detection** Working Group collaborated on a proposal "Bayesian Methods in
Syndromic Surveillance: CAR Models and Computational Implementation" submitted to the Focused Research
Groups program at NSF/DMS in September, 2006. Unfortunately, the proposal was declined, but alternative
sources of funds are being pursued. PIs on the proposal were David Banks (Duke), Gauri Datta (Georgia),
Alan Karr (NISS), James Lynch (South Carolina) and Francisco Vera (Clemson).

[1] NDHS program roles: P. Bai = NCSU graduate student, H. T. Banks = NCSU faculty participant, S.
Dediu = SAMSI postdoc, A. Y. Govan = NCSU graduate student, M. Last = NISS postdoc, A. Lloyd =
NCSU faculty fellow, H. K. Nguyen = SAMSI postdoc, M. S. Olufsen = NCSU faculty fellow and
working group leader, G. Rempala = visitor from the University of Louisville, and B. D. Slenning =
NCSU faculty participant.

References

- Bai, P., Banks, H. T., Dediu, S., Govan, A. Y., Last, M., Lloyd, A., Nguyen, H. K., Olufsen, M. S.,
  Rempala, G., and Slenning, B. D. (2007). Stochastic and deterministic models for agricultural
  production networks. *Mathematical Biosciences and Engineering*. Submitted for publication.

- Banks, H. T., Karr, A. F., Nguyen, H. K., and Samuels, Jr., J. R. (2005). Sensitivity to noise variance in
  a social network dynamics model. *Q. Appl. Math.* To appear. Available on-line at
  www.samsi.info/reports/index.shtml.

- Denogean, L. R., Karr, A. F., and Qaqish, B. F. (2006). Model-based utility of doubly random
  swapping. Technical report, National Institute of Statistical Sciences.

- Ghosh, J., Reiter, J. P., and Karr, A. F. (2006). Secure computation with horizontally partitioned data
  using adaptive regression splines. *Computational Statist. and Data Anal.* To appear.

- Hong, C.-C. and Medhin, N. G. (2006a). A nonlinear programming approach to the study of social networks. *Neural Parallel and Scientific Computation*. Submitted for publication.

- Hong, C.-C. and Medhin, N. G. (2006b). Positive and negative affinities model for social networks. *Neural Parallel and Scientific Computation*. Submitted for publication.

- Karr, A. F., Fulp, W. J., Lin, X., Reiter, J. P., Vera, F., and Young, S. S. (2006a). Secure, privacy-preserving analysis of distributed databases. *Technometrics*. To appear.

- Karr, A. F., Lin, X., Reiter, J. P., and Sanil, A. P. (2006b). Secure analysis of distributed databases. In Olwell, D., Wilson, A. G., and Wilson, G., editors, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, pages 237– 261. Springer–Verlag, New York.

- Oganian, A. and Karr, A. F. (2006). Combinations of SDC methods for microdata protection. In Domingo-Ferrer, J. and Franconi, L., editors, *Privacy in Statistical Databases 2006*, volume 4302 of *Lecture Notes in Comp. Sci.*, pages 102–113. Springer–Verlag, Berlin/Heidelberg.

- Sanil, A. P., Karr, A. F., Lin, X., and Reiter, J. P. (2004). Privacy preserving analysis of vertically partitioned data using secure matrix products. *J. Official Statist.* Submitted for publication. Available on-line at www.niss.org/dgii/technicalreports.html.

- Vance, E., Archie, E. A., and Moss, C. J. (2006). Social networks in African elephants. *J. Computational and Mathematical Organization Theory*. Submitted for publication.

- Woo, M.-J., Reiter, J. P., Oganian, A., and Karr, A. F. (2006). Global measures of data utility for microdata masked for disclosure limitation. *J. Privacy and Confidentiality*. Submitted for publication. Available on-line at www.niss.org/dgii/technicalreports.html.