

**Home:**  
804 Byrd Lane,  
Foster City, CA 94404

**Phone:**  
650.393.5190  
**Email:**  
ashish@berryconsultants.com

## ASHISH SANIL

### EMPLOYMENT

<b>Berry Consultants</b>	<i>Statistical Scientist</i>	Foster City, CA (2008 – )
<b>Bristol-Myers Squibb Company (BMS)</b>	<i>Senior Research Biostatistician</i>	Princeton, NJ (2005 – 2006, 2007-2008)
<b>Google, Inc.</b>	<i>Quantitative Analyst</i>	Mountain View, CA (2006 – 2007)
<b>National Institute of Statistical Sciences (NISS)</b>	<i>Research Statistician</i>	Research Triangle Park, NC (2000 – 2005)
	<i>Postdoctoral Fellow</i>	Research Triangle Park, NC (1998-2000)
<b>Duke University</b>	<i>Research Associate</i>	Durham, NC (1998-2000)
<b>Carnegie Mellon University (CMU)</b>	<i>Research Associate, Instructor, Teaching Assistant</i>	Pittsburgh, PA (1990-1998)

### EDUCATION

<b>Carnegie Mellon University</b>	<i>Ph.D. (Statistics)</i>	Pittsburgh, PA (1998)
	<i>M.S. (Statistics)</i>	Pittsburgh, PA (1993)
	<i>M.S. (Social &amp; Decision Sc.)</i>	Pittsburgh, PA (1993)
<b>Indian Institute of Technology</b>	<i>M.Sc. (Mathematics)</i>	Kharagpur, India (1990)
	<i>B.Sc.[honours] (Mathematics)</i>	Kharagpur, India (1988)

### PROFESSIONAL EXPERIENCE



*Statistical Scientist  
[2008–present]*

- Worked on development and packaging of a software library implementing core components for the simulation and analysis of (Bayesian) Adaptive Clinical Trials



*Senior Research Biostatistician  
[2005–2006, 2007–2008]*

- Involved in the design and software implementation for innovative Bayesian adaptive designs for optimal dose-finding clinical trials.
- Involved in the use of Bayesian methodology for dose-ranging, futility analyses, and clinical utility assessments.
- Experience with the design and analysis of a number of (Phase I) clinical trials and regulatory filing process.
- Twice awarded the BMS STAR Award (2006, 2007)



*Quantitative Analyst  
[2006–2007]*

Collaborated with a multidisciplinary team of engineers and analysts on a wide range of problems. Applied statistical methods to the problems encountered in the evaluation and optimization of Google's ads quality.

## NISS

(National Institute of Statistical Sciences)

*Research Statistician [2000–2005]  
Postdoctoral Fellow (concurrent with  
Duke University Research Associate position) [1998–  
2000]*

NISS is a research organization that undertakes cross-disciplinary projects involving the statistical sciences. At NISS:

- Developed, implemented, published and presented new methodology.
- Identified and explored new research directions and opportunities, and participated in writing grant proposals.
- Mentored postdoctoral fellows, and oversaw software development efforts.

### Selected projects:

- **Software Engineering:** Models for predicting software failure, and optimal instrumentation
- **Prediction models for automobile sales data** (a data mining project with sales data of a major US automobile manufacturer).
- **Modeling Web-browsing behavior:** formulated Bayesian models for the patterns of users' traversal of websites.
- **Modeling travel demand:** developed and evaluated classification models to predict household-level travel demand based on property-tax data.
- **Statistical Disclosure Limitation:** Developed new methods and algorithms for a number of problems in this area (that is addresses the problem of disseminating statistically useful data without jeopardizing the confidentiality of data subjects).

### Selected projects:

- Analysis of multivariate time series data: Conducted an analysis of a multivariate time series of manufacturing process data to identify the factors contributing to disruptions in a fiberglass manufacturing process.
- Development of software tools for analysis of supermarket data: Helped in the setup, and developed tools/scripts for data analysis and extraction from large data sets of scanner records from a retail grocery chain.
- Groundwater pollution study: Modeled the spread of the contaminants in groundwater.

### SELECTED TEACHING EXPERIENCE

- |   |              |
|---|--------------|
| • Introductory Probability and Statistical Inference for the Computational Finance program [Instructor] | CMU, 1996-97 |
| • Linear Models and Time Series Analysis courses for the Computational Finance program [TA]             | CMU, 1995-96 |
| • Undergraduate Mathematical Probability and Statistics courses [Instructor]                            | CMU, 1995-96 |
| • Information Systems courses (Structured Analysis and Design, Database Management Systems) [Head TA]   | CMU, 1990-92 |

### GRANTS

NSF Grant ITR-0218759 (co-PI with Alan Karr): *Bayesian Models Linking Web Site Structure and Usage*.

### PATENTS

- *Market Maturity Optimization Of Web-based Service*: Ashish Sanil, Diane L. Tang, Nicholas C. Fox, and Avichal Garg.
- *Methods, Systems and Computer Program Products for Identifying Pharmacophores in Molecules Using Inferred Conformations and Inferred Feature Importance*: Jun Feng, Ashish P. Sanil and S. Stanley Young.

## PUBLICATIONS

### Doctoral Dissertation

An Approach For Selection Of Nonparametric Regression Methods (under the supervision of Prof. William Eddy). The thesis addresses the issue of evaluating the generalizability of functional relationships between predictor attributes and a response attribute as learned from data (for a certain class of learning algorithms that includes generalized smoothing splines, radial basis functions, etc.) by unifying them under a Bayesian framework.

### Reviews

- Sanil, A. (2003). Review of “Principles of Data Mining” by Hand, Mannila and Smyth. *Journal of the American Statistical Association*, **98**, No. 461, March 2003.

### Refereed Journals and Book Chapters

- Karr, A., Lin, X., Reiter, J. Sanil, A. (2008). Privacy Preserving Analysis of Vertically Partitioned Data Using Secure Matrix Products, *Journal of Official Statistics* (to appear).
- Haran, A., Karr, A., Last, M., Orso, A., Porter, A., Sanil, A. and Fouche, S. (2007). Techniques for Classifying Executions of Developed Software to Support Software Engineering Tasks, *IEEE Trans. Software Engineering*, May 2007, 287-304.
- Karr, A., Lin, X., Reiter, J. Sanil, A. (2006). Secure Regression on Distributed Databases, *Statistical Methods in Counterterrorism*. by D. Olwell and A. G. Wilson (eds.), ASA-SIAM Series on Statistics and Applied Probability.
- A. F. Karr, C. N. Kohnen, A. Oganian, J. P. Reiter, A. P. Sanil (2006). A framework for evaluating the utility of data altered to protect confidentiality. *The American Statistician*. **60**, No. 3.
- Jun Feng, Ashish Sanil, and S. Stanley Young (2006). PharmID: Pharmacophore Identification Using Gibbs Sampling. *Journal of Chemical Informatics and Modeling*, Feb 2006 (web version).
- Gomatam, S., Karr, A., Sanil, A. (2005) Data Swapping as a Decision Problem, *Journal of Official Statistics*, **21**, No. 4.
- A. F. Karr, J. Feng, X. Lin, J. P. Reiter, A. P. Sanil and S. S. Young (2005). Secure analysis of distributed chemical databases without data integration. *Journal of Computer-Aided Molecular Design*, November 2005.
- Gomatam, S., Karr, A., Reiter, J. Sanil, A. (2005). Data Dissemination and Disclosure Limitation in a World Without Microdata: A Risk-Utility Framework for Remote Access Analysis Servers, *Statistical Science*, **20** , No. 2.
- Karr, A., Lin, X., Reiter, J. Sanil, A. (2005). Secure Regression on Distributed Databases, *Journal of Computational and Graphical Statistics*, **14**, no. 2.
- Karr, A., Lin, X., Reiter, J. Sanil, A. (2004). Analysis of Integrated Data Without Data Integration, *Chance*, **17**, No. 3, Summer 2004.
- Dobra, A., Karr, A., and Sanil, A. (2003). Preserving Confidentiality of High-dimensional Tabulated Data: Statistical and Computational Issues. *Statistics and Computing*, **13**, No. 4, October 2003.
- Sanil, A., Gomatam, S. and Karr, A. (2003). NISS WebSwap: A Web-Service for Data Swapping. *Journal of Statistical Software*, **8**, No. 7.
- Dobra, A., Karr, A., and Sanil, A. (2003). Table Servers: Protecting Confidentiality in Tabular Data Releases. *Communications of the ACM*, January 2003.
- Dobra, A., Fienberg, S., Karr, A., and Sanil, A. (2002). Software Systems for Tabular Data Releases. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **10**, No. 5: 529-544.

- Lee, J., Holloman, C., Karr, A., and Sanil, A. (2002). Analysis of Aggregated Data in Survey Sampling with Application to Fertilizer/Pesticide Usage Survey. *Research in Official Statistics*, **4**, 101-106.
- Karr, A., Lee, J. and Sanil, A. (2001). Web-based Systems that Disseminate Information from Data but Protect Confidentiality, in *Advances in Digital Government* by Ahmed K. Elmagarmid and William J. McIver Jr.(eds.), Kluwer Academic.
- Karr, A., Lee, J., Sanil, A., Hernandez, J., Karimi, S., and Litwin, K. (2001). Disseminating Information but Protecting Confidentiality, *IEEE Computer*, February 200, 36-37.
- Fienberg, S., Makov, U. and Sanil, A. (1997). A Bayesian Approach to Data Disclosure: Optimal Intruder Behavior for Continuous Data, *Journal of Official Statistics*, **13**, No. 1: 75-89.
- Sanil, A., Banks, D. and Carley, K. (1995). Models for Evolving Fixed-Node Networks: Model Fitting and Model Testing. *Social Networks*, **17**, No. 5: 529-544.

### **Selected Conference Papers**

- Sanil, A., Karr, A., Lin, X., Reiter, J. (2004). Privacy Preserving Regression Modeling via Distributed Computation, *KDD-2004*
- Haran, M., Karr, A., Sanil, A. (2004). A Model for Relating Browsing Behavior to Site Design on the World Wide Web, *Proceedings of the Joint Statistical Meetings, 2004*.
- Gomatam, S., Karr, A., Liu, C. and Sanil, A. (2003). Data Swapping: A Risk-Utility Framework. *dg.02003 (Digital Government Conference)*
- Sohn, M., Bromhal, G., Small, M., Pantazidou, M., Sanil, A. and Eddy, W. (1997). Updating Uncertainty in Groundwater Contamination. *Proceedings of the Water Resources Planning and Management Conference of the American Society of Civil Engineers*.

### Selected Presentations

- Issues in 'Real Data' Mining. Presented as an invited talk at *Interface 2004: Computational Biology and BioInformatics*, 36<sup>th</sup> Symposium on the Interface (of Computing Science and Statistics). May 26-29, 2004. Baltimore, MD. [Also presented as invited talk at *2004 Quality and Productivity Research Conference*.]
- Pointers from Research on Data Confidentiality and Data Quality (in Homeland Security and Related Issues). Invited talk at *Interface 2003: Security and Infrastructure Protection*, 35<sup>th</sup> Symposium on the Interface. March 12-15, 2003. Salt Lake City, Utah.
- Statistical Utility of Disclosure Limited Releases of Cross-Tabulated Data Presentation at the *Joint Statistical Meetings* (of the American Statistical Association, The International Biometric Society (ENAR and WNAR), the Institute of Mathematical Statistics, and the Statistical Society of Canada). August 3-7, 2003, San Francisco, CA.
- A Simple, Statistical Approach to Data Quality. Presentation at the *Joint Statistical Meetings*. August 11-15, 2002. New York, NY.
- Geographic Aggregation Procedures for Data Disclosure Limitation. Invited talk at *Interface 2000: Computing Science and Statistics*, 32<sup>nd</sup> Symposium on the Interface. April 5-8, 2000. New Orleans, LA.
- Web-Based Query Systems for Risk-Limited Dissemination of Confidential Data. Invited talk at the *AT&T Research Labs*. October 3, 2000, Florham Park, NJ. [Also presented as invited talk at *Joint Statistical Meetings* (2000), *The Sixth Great Lakes Symposium on Applied Statistics* (2000), *International Biometric Society Spring Meetings* (2001), and *The Spring Research Conference on Statistics in Industry and Technology* (2001).]
- Several presentations and briefings on Disclosure-Limited Dissemination of Data and Data Quality at the US Census Bureau, National Agricultural Statistics Service, Bureau of Labor Statistics, National Center for Education Statistics, and Bureau of Transportation Statistics.