

Norman S. Matloff

Norman S. Matloff

Professor of Computer Science
University of California at Davis
Davis, CA 95616
(530) 752-1953
nsmatloff@ucdavis.edu

June 17, 2024

Background

Born and raised in Los Angeles.

Ph.D. in (Theoretical) Mathematics, UCLA.

Significant Honors and Accomplishments

- Eric Ziegel Award, 2018
- 6 books, 4 book chapters, numerous research papers
- Pioneering work in causal inference (ATE), data privacy, parallel computation for data science
- Keynote speaker, *SatRdays*, April 2019, UCLA
- Keynote speaker, *R/Finance 2018*, June 2018, Chicago
- Keynote speaker, *useR! 2017*, July 2017, Brussels, Belgium
- Editorial Board, *Journal of Statistical Software*, 2012-
- Editor in Chief, *The R Journal* (publication of the R Foundation), 2017-2018
- Executive Editor (one of three), *The R Journal* 2016-2018
- Past appointee to IFIP Working Group 11.3 on Database Security (UNESCO-sponsored international research and policy committee)
- UCD Distinguished Teaching Award (campuswide), 2006
- UCD Distinguished Public Service Award (campuswide), 2002
- UCD Outstanding Faculty Advisor Award (campuswide), 1994
- UCD Computer Science Club's Professor of the Year Award, 1995, 2002, 2005 and 2011

Research/Writings

dsld: A Socially Relevant Tool for Teaching Statistics, Taha Abdullah, Arjun Ashok, Brandon Estrada, Norman Matloff, Aditya Mittal, arXiv, 2024

TowerDebias: A Novel Debiasing Method based on the Tower Property, Norman Matloff and Aditya Mittal arXiv, 2024

An S-Curve Method for Abrupt and Gradual Changepoint Analysis, with L. Jiang and C. Kennedy, the *Journal of Data Science*, in press

RWN: A Novel Neighborhood-Based Method for Statistical Disclosure Control, with Noah Perry and Patrick Tendick, *Transactions on Data Privacy*, 2024

The Art of Machine Learning: Algorithms+Data+R , NSP, June 2023

An S-Curve Method for Abrupt and Gradual Changepoint Analysis, with L. Jiang and C. Kennedy, SDSS 2023

Belated Tribute to a Great Mathematician, and Some Related Personal Thoughts on PhD Education, *Celebratio Mathematica*, 2022

Walk a Mile in Their Shoes: a New Fairness Criterion for Machine Learning, arXiv, 2022

K-Means Clustering Usage in Datasets with Missing Values, with WX Zhang, useR! 2022

Modernizing k-Nearest Neighbor Software, with Robin Elizabeth Yancey and Bochao Xin, Stat, 2021

TDAsweep: A Novel Dimensionality Reduction Method for Image Classification Tasks, with Yu-Shih Chen and Melissa Goh, arXiv, 2020

R vs. Python for Data Science, **invited** paper, SDSS 2020

Modernizing k-Nearest Neighbor Software, with Robin Elizabeth Yancey and Bochao Xin, SDSS 2020

prVis: a Novel Method for Visual Dimension Reduction, with Tiffany Jiang, Robert Tucker and Allan Zhao, useR! 2019

Probability and Statistics for Data Science: Math + R + Data , CRC, 2019

Methods for Visualizing Dimension Reduction in R, with Tiffany Jiang, Robert Tucker and Allan Zhao, SDSS, 2019

Extension of the Tower Method for Missing Values to Time Series, with Pete Mohanty, R/Finance 2019, **keynote address**

Fast, General Parallel Computation for Machine Learning, P2PS workshop, with Robin Yancey, ICPP 2018

Fast Computation of Large-Scale Mixed Effects Models, with Robin Yancey, Proceedings of the JSM 2018

revisit: a Statistical Teaching Tool in R, with Emily Watkins and Tiffany Chen, JSM 2018

The R Language: a Powerful Tool for Taming Big Data, **invited** book chapter with Robin Yancey and C. Fitzgerald, *Encyclopedia of Big Data Technologies*, Springer, 2018

The Data Privacy Problem: Computer Science, Statistics and Future Directions, **invited paper**, SAE, July 2017, Paris, France

Statistical Regression and Classification: from Linear Models to Machine Learning (professional monograph/textbook), Chapman and Hall, 2017; **Eric Ziegel Award, Best Reviewed Book of 2017**, *Technometrics*

A New Framework for Random Effects Models, JSM 2016.

Statistics and R for Analysis of Elimination Tournaments, Ariel Shin and Norman Matloff, *userR!* 2016.

Rectools: an Advanced Recommender System, Pooja Rajkumar and Norman Matloff, *userR!* 2016.

Software Alchemy: Turning Complex Statistical Computations into Embarrassingly-Parallel Ones, Norman Matloff, *Journal of Statistical Software*, 71 (2016)

A Closer Look at What We Should Mean by "Big" in "Big Data," *Handbook of Big Data* (**invited chapter** in professional monograph), Hans Buhlmann *et al* (eds.), Chapman and Hall, 2016

Improved Estimation of Class Probabilities through Unlabeled Data , arXiv:1510.01422, 2016

A New Method for Avoiding Data Disclosure While Automatically Preserving Multivariate Relations, N. Matloff and P. Tendick, arXiv:1510.04406, 2016

A Different Approach to the Problem of Missing Data, with Xiao (Max) Gu, *Proceedings of the Joint Statistical Meetings*, 2015

Parallel Computation for Data Science (professional monograph), Chapman and Hall, 2015

A New Approach to the Parallel Coordinates Method for Large Data Sets, with Yingkang Xie, *Proceedings of the Joint Statistical Meetings*, 2014

A Package for Parallel Matrix Powers, with Jack Norman, *userR! 2014*

Regression Fit Diagnostics Using *freqparcoord*, with Yingkang Xie, *userR! 2014*

Long Live (Big Data-fied) Statistics!, **invited paper**, *Proceedings of JSM 2013*, 98-108

Immigration and the Tech Industry: As a Labour Shortage Remedy, for Innovation, or for Cost Savings?, *Migration Letters* 2013 (extension of **invited** paper presented at Migration and Competitiveness: Japan and the US, Berkeley/Japan, 2012, funded by Sloan Foundation)

Are Foreign Students the 'Best and Brightest'? Data and implications for immigration policy, **invited paper**, Economic Policy Institute, 2013.

Efficient Parallel R Loops on Long-Latency Platforms, **invited paper**, *Proceedings of Interface 2012-- Future of Statistical Computing: Internet Scale Data, Flexible Modeling, and Visualization*, Scott, Wickham and Morris (eds.), 2012.

Parallel R Revisited (**invited paper**, abstract), *useR! Abstract Booklet, 2012*, Vanderbilt University, http://biostat.mc.vanderbilt.edu/wiki/Main/UseR-2012#Abstract_Booklet

Debate: H-1B Visa (**invited book chapter**), in *Debates on U.S. Immigration*, Judith Gans, Elaine Replogle, Daniel Tichenor (eds.), Sage Publications, 2012

The Art of R Programming (professional book), No Starch Press, 2011.

Rdsm: Distributed (Quasi-)Threads Programming in R (abstract), *useR! 2010*, Gaithersburg, MD, http://www.r-project.org/conferences/useR-2010/abstracts/_Abstracts.pdf

Programming on Parallel Machines (open textbook), 2010-.

KarmaNET: Leveraging Trusted Social Paths to Create Judicious Forwarders (with X. Lu, M. Spear and F. Wu), First International Conference on Future Information Networks, 2009, pp.218-223.

The Art of Debugging, with GDB, DDD and Eclipse (professional book), No Starch Press, with Peter Salzman, 2008.

Virtual Migration, by A. Aneesh (**invited** book review), by N. Matloff, *Journal of International Migration and Integration*, December 2008, 425-427.

A New Method for Rule Finding Via Bootstrapped Confidence Intervals. *SIAM Conference on Data Mining*, 2008, 547-552.

Using Soft-line Recursive Response to Improve Query Aggregation in Wireless Sensor Networks (with X. Lu, M. Spear, K. Levitt, F. Wu). *2008 IEEE International Conference on Communications*, 2309-2316.

A Synchronization Attack and Defense in Energy-Efficient Listen-Sleep Slotted MAC Protocols (with X. Lu, M. Spear, K. Levitt, F. Wu), *SECURWARE archive Proceedings of the 2008 Second International Conference on Emerging Security Information, Systems and Technologies*, 403-411.

Availability-Aware Provisioning Strategies for Differentiated Protection Services in Wavelength-Convertible WDM Mesh Networks (with B. Mukherjee, H. Zang, J. Zhang and K. Zhu). *IEEE/ACM Transactions on Networking*, 2007, 15, 5, 1177-1190.

Revisiting the Issue of Performance Enhancement of Discrete Event Simulation Software (with Alex Bahouth, Steven Crites, and Todd Williamson), *Proceedings of the 40th Annual Simulation Symposium*, 2007, 114-122.

From Algorithms to Z-Scores: Probabilistic and Statistical Modeling in Computer Science (open textbook), 2006-.

On the Adverse Impact of Work Visa Programs on Older U.S. Engineers and Programmers (**invited** paper). *California Labor and Employment Law Review* (a publication of the California State Bar Association), August 2006.

Estimation of Internet File-Access/Modification Rates, *ACM Transactions on Modeling and Computer Simulation*, 2005, 15, 3, 233-253.

Offshoring: What Can Go Wrong? (**invited** article). *IEEE IT Pro*, July/August 2005, 39-45.

A Careful Look at the Use of Statistical Methodology in Data Mining. In T.Y. Lin, Wesley Chu and L. Matzlack (eds.), *Foundations of Data Mining and Granular Computing*, Springer-Verlag Lecture Notes in Computer Science, 2005, 101-117.

Globalization and the American IT Worker (**invited** column). *Communications of the ACM*, November 2004, 27-29.

On the Need for Reform of the H-1B Non-immigrant Work Visa in Computer-Related Occupations (**invited** paper). *University of Michigan Journal of Law Reform*, Fall 2003, Vol. 36, Issue 4, 815-914 (99 pages).

Toward a Statistical Foundation for Data Mining. *ICDM '02 Workshop on Foundations of Data Mining and Knowledge Discovery*, 2002, 125-130.

PerlDSM: A Distributed Shared Memory System for Perl. *Proceedings of PDPTA 2002*, 2002, 63-68.

I-Tuples: A Programmer-Controllable Performance Enhancement for the Linda Environment. *Proceedings of PDPTA 2001*, 2001, 357-361.

Decentralized Task Assignment Is Scalable. *Proceedings of PDPTA 2000*, 2000, 285-289.

The Effectiveness of the Programmed Backoff Method in the Presence of Background Traffic. *Proceedings of PDPTA 99*, 1999, 2332-2336.

TupleDSM: An Educational Tool for Software Distributed Shared Memory. *Proceedings of the Workshop on Computer Architecture Education*, Orlando, Florida, January 1999.

Analysis of a Programmed Backoff Method for Parallel Processing on Ethernets. *Network-Based Parallel Computing*, Dhabaleswar Panda and Craig Stunkel, eds., Lecture Notes in Computer Science 1362, Springer-Verlag, 1998, pp. 110-117.

Network-Specific Performance Enhancements for PVM (with Gregory Davies). *Proceedings of the Fourth IEEE International Symposium on High-Performance Distributed Computing*, 1995, 205-210.

KuaiXue: A Computer Tool for Teaching Chinese. *Proceedings of the International Conference of New Technologies in Teaching and Learning Chinese*, Burlingame, California, May 1995 (extended abstract).

- A Modified Random Perturbation Method for Database Security (with Patrick Tendick). *ACM Transactions on Database Systems*, 1994, 19(1), 47-63.
- A Locally Cache-Coherent Multiprocessor Architecture (with K. Rich). *Proceedings of SS'93 High Performance Computing*, Calgary, Canada, June 1993, 411-418.
- A Probabilistic Limit on the Virtual Size of Replicated Disk Systems (with R. Lo). *IEEE Transactions on Knowledge and Data Engineering*, 1992, 4(1), 99-102.
- An Argument Against Scalable Cache Coherency, *Computer Architecture News*, 1991, 19, 4, 117-123.
- Simulation Event-List Algorithms for Use in Virtual Memory Systems (with C. Martel and D. Naor). *Computer Journal*, 34, 5, 1991, 428-437.
- IBM Microcomputer Architecture and Assembly Language: A Look Under the Hood* (textbook), Prentice-Hall, 1991.
- Dynamic Control and Accuracy of the p_1 -Persistent Protocol Using Channel Feedback (with B. Mukherjee, A. Lantz and S. Bannerjee). *IEEE Transactions on Communications*, 1991, 39(6), 887-898.
- Statistical Hypothesis Testing: Problems and Alternatives. *Journal of Economic Entomology*, 20, 1991, 1246-1250.
- A "Greedy" Approach to the Write Problem in Shadowed Disk Systems (with R. Lo). *Proceedings of the Sixth International Conference on Data Engineering*, 1990, 553-558.
- Selectivity Estimation Using Homogeneity Measurement (with M. Chen and L. McNamee). *Proceedings of the Sixth International Conference on Data Engineering*, 1990, 304-310.
- Estimating a Mixing Distribution in a Multiple Observation Setting (with Y.P. Mack). *Statistics and Probability Letters*, 1990, 369-376.
- On the Value of Predictive Information in a Scheduling Problem. *Performance Evaluation*, 1989, 309-315.
- Fixed Optical Interconnects for Concurrent Computer Systems (with C. Eldering, S. Kowel, R. Brinkley, T. Schubert and R. Gosula). *Proceedings of the 1989 SPIE International Symposium on Optical and Optoelectronic Applied Science and Engineering*, 72-82.
- Dynamic Control of the p_1 -Persistent Protocol Using Channel Feedback (with B. Mukherjee, A. Lantz and M. Moh). *Proceedings of INFOCOM 89*, 858-865.
- Performance Analysis of the OPTIMUL Multiprocessor Interconnect (with T. Schubert, S. Kowel, C. Eldering). *Proceedings of the International Phoenix Conference on Computers and Communications*, 1989, 60-63.

The “Curse of Dimensionality” in Database Security (with P. Tendick) *DATABASE SECURITY, II: Status and Prospectus*, C. Landwehr (ed.), North-Holland Publishers, 225-232, 1989.

Electro-Optical Interface (with S. Kowel and C. Eldering) U.S. Patent No. 4,813,772, March 21, 1989.

OPTIMUL: An Optical Interconnect for Multiprocessor Systems (with Stephen Kowel and Charles Eldering). *Proceedings of the ACM International Conference on Supercomputing*, 16-24, St. Malo, France, 1988.

A Predictive Flow Control Policy for ISDN Token Rings (with Georgia Fuller and Patrick Tendick). *Proceedings of the 1988 Computer Networking Symposium*, Washington, D.C., 1988, 130-134.

Inference Control Via Query Restriction Vs. Data Modification: A Perspective. In *DATABASE SECURITY: Status and Prospectus*, C. Landwehr (ed.), North-Holland Publishers, 159-166, 1988.

Recent Results on the Noise Addition Method for Database Security (with Patrick Tendick). *Proceedings of the Joint ASA-IMS Meetings*, 1987, 406-409.

Probability Modeling and Computer Simulation, Applied to Engineering and Computer Science (textbook). PWS-Kent, 1988.

Distribution-Free Analysis of Repairable Fault-Tolerant Systems. *Microelectronics and Reliability*, 1987, 27, 549-556.

A Multiple Disk System for Both Fault Tolerance and Improved Performance. *IEEE Transactions on Reliability* (special issue on Fault-Tolerant Computing Techniques and Systems), R-36, 1987, 199-201.

Future Applications of Ordered Polymeric Thin Films (with S. Kowel, R. Selfridge, C. Eldering, P. Stroeve, B. Higgins, M. Srinivasan and L. Coleman). *Thin Solid Films*, 152, 377-403.

Pipelined Designs for Binary Search Processors. *Proceedings of the Twentieth Annual Asilomar Conference on Signals, Systems and Computers*, 1986, 532-535.

Another Look at the Use of Noise Addition for Database Security. *Proceedings of the 1986 IEEE Symposium on Security and Privacy*, April 1986, pp. 173-180.

Statistical Hypothesis Testing in Biology: A Contradiction in Terms (with D. Jones). *Journal of Economic Entomology*, 1986, 1156-1160.

Dynamic Determination of Disk Reorganization Times (with C. Wang). *Proceedings of the Eighteenth Annual Asilomar Conference on Circuits, Systems and Computers*, 1984, 337-340.

The Asymptotic Distribution of an Estimator of the Bayes Error Rate (with R. Pruitt). *Pattern Recognition Letters*, 1984, 2, 271-274.

Update Methods for Reduction of Run Time in Simulation Studies. *Computational Statistics and Data Analysis*, 1984, 237-242.

Use of Lattice Structures for Reduction of Simulation Run Time. *Proceedings of the 1984 Winter Simulation Conference*, 1984, 124-126.

A Reliable Method of Determining the Rank of a Matrix (with C. Wang). *Proceedings of the 1984 ACM/SIGNUM Conference on Numerical Computations and Mathematical Software for Microcomputers* (April 1984 issue of ACM/SIGNUM Newsletter), 1984, 19.

Efficient Methods for Microcomputer Simulation Software. *Proceedings of the 1984 International Symposium on Mini and Microcomputers and Their Applications*, 16-18.

A Comparison of Two Methods for Estimating Optimal Weights in Regression Analysis (with R. Rose and R. Tai). *Journal of Statistical Computation and Simulation*, 1984, 19, 265-274.

Use of Covariates in Randomized Response Settings. *Probability and Statistics Letters*, 1984, 2, 31-34.

James-Stein Estimation in a Prediction Context. *Communications in Statistics: Simulation and Computation*, 1982, B11, 589-601.

Use of Regression Functions for Improved Estimation of Means. *Biometrika*, 1981, 68, 685-689.

An Alternative to the Analysis of Covariance for Comparing Randomized Treatments. *Communications in Statistics: Theory and Methods*, 1981, A10, 2015-2024.

The Jackknife. *Applied Statistics* (Algorithms Section), 1980, 29, 115-117.

A Dissonant Voting Model: Nonergodic Case. *Zeitschrift für Wahrscheinlichkeitstheorie*, 1980, 51, 63-78.

Ergodicity Conditions for a Dissonant Voting Model. *Annals of Probability*, 1977, 5, 371-386.

Public Service

Have written frequently about minority and social issues, especially on age discrimination in the software industry.

Have presented invited testimony on a number of occasions to the U.S. Senate and House of Representatives. Have served as a consultant to the U.S. Dept. of Commerce and Dept. of Health and Human Services during the Clinton administration.

Frequently serve as an invited panelist on computer industry hiring practices, in forums sponsored by industry, academia, government and public-affairs groups,

such as Migration and Competitiveness: Japan and the United States, the Stanford University Computer Project Conference, the Boston University Workshop on Migration of Foreign Scientists and Engineers to the United States, the ITAA/Dept. of Commerce Convocation, the Commonwealth Club of San Francisco, the Gartner Group Application Development Summit, MEPTECH, Silicon Valley Power Breakfast, Software Development Expo, the California Governor's Older Worker and Exemplary Employer Conference, etc. His recommendations on careers in the computer field has been occasionally sought by writers of career-advice columns, and syndicated columnist Joyce Lain Kennedy features Dr. Matloff's e-newsletter on career issues in her books, *Resumes for Dummies*, *Job Hunting for Dummies* and *Cover Letters for Dummies*.

Have written articles (in many cases by invitation of the magazine or newspaper) for the *New York Times*, the *Washington Post*, the *Public Interest*, the *New Democrat* (a publication of the Democratic Leadership Council), the *National Review*, the *Los Angeles Times*, the *San Francisco Chronicle*, *AsianWeek*, the *San Jose Mercury News*, *Controller Quarterly* (published by the California State Controller's Office) and so on.

Have been interviewed or cited by NBC, ABC, CBS, CNN, NPR, PBS, the Voice of America, the *New York Times*, the *Los Angeles Times*, the *Washington Post*, the *Chicago Tribune*, the *San Francisco Chronicle*, the *San Jose Mercury News*, the *Boston Globe*, the *San Diego Union-Tribune*, the *Dallas Morning News*, the Associated Press, *Mother Jones*, the *New Republic*, the *Wall Street Journal*, *Investor's Business Daily*, *Business Week*, *US News and World Report*, *Science Magazine*, *Computerworld*, the *Electronic Engineering Times*, *Upside*, *AsianWeek*, *India West* and many others.

File translated from T_EX by T_TH, version 2.88.
On 20 Jul 2002, 11:47.