A Bibliography of Algorithm Experimentation

Catherine C. McGeoch

This annotated bibliography lists sources that address methodology, tools and techniques, and standards for experimental research on algorithms. I've tried, in the annotation, to summarize both the content and the technical level. No attempt has been made to catalog the huge number of published *examples* of experimental and computational research on algorithms, although some general publication venues are listed.

We begin with a shortlist of resources for the beginning experimentor. These sources, taken together, provide good overall coverage of the main issues of experimental research, at about the graduate student level. The reference list that follows contains works on more specialized, advanced, and technical topics.

A Beginner's Reading List.

- For general advice, tips, and guidelines on algorithmic experimental resaerch, read Barr et al. [**BGK$^+$95**], McGeoch [**McG96b**], [**McG00**], Johnson [**Joh01**], and Moret [**Mor01**].
- Bentley [**Ben91**] shows how to build a Unix environment that supports efficient testing and data collection. Some remarks on this subject may also be found in [**McG01b**].
- Ahuja, Magnanti, and Orlin [**AMO93**] discuss algorithmic performance measures that closely predict running times and that (to some extent) can be extrapolated across machine environments.
- For an introduction to statistical concepts and techniques that are most relevant to experimental research on algorithms, read the book by Cohen [**Coh95**] and/or the article by Coffin and Saltzman [**CS00**]. Both of these sources contain examples and illustrations drawn from algorithmic research problems. For a readible and authoritative discussion of techniques of graphical data analysis, see Chambers et al. [**CCKT83**].
- Crowder, Dembo, Mulvey [**CDM78**], and the follow-up article by Jackson et al. [**JBNP91**] give criteria and standards for presentation of experimental research in published work. The first provides a handy checklist for reviewers. See also Johnson [**Joh01**] for advice on what kind of information to include in your talk or paper. McGeoch and Moret [**MM99**] give guidelines for presenting talks on experimental research on algorithms.

References

[AMO93]   R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows: Theory, algorithms, and applications*, Prentice Hall, Englewood Cliffs, NJ, 1993, •Chapter 18 contains good advice on selecting appropriate cost measures.

[And01]   R. Anderson, *The role of experiment in the theory of algorithms*, this volume, 2001.

[Bar87]   R. R. Barton, *Testing strategies for simulation optimization*, Proceedings of the 1987 Winter Simulation Conference (A. Thesen, H. Grant, and W. D. Kelton, eds.), Society for Computer Simulation, 1987, pp. 391–401.

[BB98]   R. Battiti and A. Bertossi (eds.), *Alex'98: Workshop on algorithms and experiments*, *rtm.science.unitn.it/alex98/*, 1998, •Submitted and invited papers by workshop participants.

[Ben82]   J. L. Bentley, *Writing efficient programs*, Prentice Hall, Englewood Cliffs, NJ, 1982, •Describes a systematic methodology for reducing constant factors in implementations.

[Ben84]   J. L. Bentley, *A case study in applied algorithm design*, Computer **17** (1984), no. 2, 75–88, •TSP algorithms applied to a problem in pen-plotter scheduling.

[Ben90]   J. L. Bentley, *Experiments on traveling salesman heuristics*, Proceedings of the First ACM-SIAM Symposium on Discrete Algorithms, ACM-SIAM, New York, Philadelphia, 1990, •Discusses both the research problem and the experimental methodology., pp. 91–99.

[Ben91]   J. L. Bentley, *Tools for experiments on algorithms*, CMU Computer Science: A 25th Anniversary Commemorative (R. F. Rashid, ed.), ACM Press, New York, 1991, •How to build Unix tools to make an efficient and reliable experimentation environment. Check out the 'Ten Commandments of Experimentation on Algorithms.'.

[Ben99]   J. L. Bentley, *Programming pearls, second edition*, Addison Wesley, Reading, MA, 1999, •A must-read for algorithm engineers. Part II considers both asymptotic and constant-factor performance issues.

[BFS83]   P. Bratley, B. L. Fox, and L. E. Schrage, *A guide to simulation*, Springer-Verlag, New York, 1983, •A textbook aimed at the first-year graduate level. Covers statistical techniques as well as implementation issues, with attention to practical concerns.

[BGK+95]   R. S. Barr, B. L. Golden, J. P. Kelly, M. G. C. Resende, and W. R. Stewart, *Designing and reporting on computational experiments with heuristic methods*, Journal of Heuristics **1** (1995), no. 1, 9—32, •Advice and standards for experimental research on heuristic algorithms.

[BS01]   A. Buchsbaum and J. Snoeyink, *ALENEX01: Workshop on Algorithm Engineering and Experiments*, *www.research.att.com/~alb/ALENEX01/*, Washington, D.C., January 2001, •Third in an annual series of workshops. The website has links to previous meetings.

[CCKT83]   J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey, *Graphical methods for data analysis*, Duxbury Press, Boston, 1983, •Readible and authoratative reference on graphical analysis techniques.

[CDM78]   H. P. Crowder, R. S. Dembo, and J. M. Mulvey, *Reporting computational experiments in Mathematical Programming*, Mathematical Programming **15** (1978), 316–329, •This classic paper stands the test of time. The authors present criteria for reporting experimental results and provide a handy checklist for reviewers.

[Che82]   R. C. H. Cheng, *The use of antithetic variates in computer simulation*, Journal of the Operations Research Society **33** (1982), 229—237, •Illustrates the application of a variance reduction technique.

[Cle85]   W. S. Cleveland, *Elements of graphing data*, Wadsworth, Monteray, CA, 1985, •A good general reference covering both analysis and presentation.

[Cle93]   W. S. Cleveland, *Visualizing data*, Hobart Press, Summit, NJ, 1993, •Graphical and visualization tools for getting better views of data.

[Coh95]   P. R. Cohen, *Empirical methods for artificial intelligence*, MIT Press, Cambridge, 1995, •Statistical methods and practical advice for algorithm experimentation. This textbook is relevant to all kinds of algorithmic research, not just heuristics in Artificial Intelligence.

[CS00]   M. Coffin and M. J. Saltzman, *Statistical analysis of computational tests of algo-rithms and heuristics*, INFORMS Journal on Computing **12** (2000), no. 1, 24–44, Introduction and survey of statistical techniques most relevant to algorithmic exper-imentation.

[Dev86]   L. Devroye, *Non-uniform random variate generation*, Springer-Verlag, New York, 1986, •Algorithms, code, theory: a must-have reference.

[DG87]   J. J. Dongarra and E. Grosse, *Distribution of mathematical software via electronic mail*, Communications of the ACM **30** (1987), no. 5, 403–407, Available through the mail server netlib@research.att.com. •The NETLIB collection of instances for network problems.

[FMS00]   R. Fleischer, B. M. E. Moret, and E. M. Schmidt (eds.), *Dagstuhl seminar on ex-perimental algorithmics*, Schloss Dagstuhl Seminar No. 00371, September 2000, Ab-stracts: *www.dagstuhl.de/DATA/Seminars/00/.* Full papers to appear in *Lecture Notes in Computer Science*, Springer Verlag, 2001.

[FR92]   G. S. Fishman and D. S. Rubin, *Bounding the variance in Monte Carlo experiments*, Operations Research Letters **11** (1992), 243–248, •How to use variance estimations to decide how many tests are needed.

[FY97]   G. S. Fishman and L. S. Yarberry, *An implementation of the batch means method*, INFORMS Journal on Computing **9** (1997), no. 3, 296–310, •Illustrates a variance reduction technique.

[GAWB86]   B. L. Golden, A. A. Assad, E. A. Wasil, and E. Baker, *Experimentation in optimiza-tion*, European Journal of Operational Research **27** (1986), 1–16, •General advice and tips on experimentation.

[GM99a]   A. V. Goldberg and B. M. E. Moret (eds.), *CATS: Combinatorial Algorithms Test Sets*, *www.jea.acm.org/CATS/*, since 1999, •A repository for input sets for com-binatorial algorithms. Contains test sets for several graph problems and NP-Hard problems, with links to related sites.

[GM99b]   M. C. Goodrich and C. C. McGeoch (eds.), *Workshop on algorithm engineering and experimentation (alenex99)*, Lecture Notes in Computer Science, no. 1619, Springer Verlag, 1999.

[Gre90]   H. Greenberg, *Computational testing: Why, how, and how much*, ORSA Journal on Computing **2** (1990), 94–97, •A short article on larger issues of computational testing.

[GS85]   B. L. Golden and W. R. Stewart, *Empirical analysis of heuristics*, The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization (E. Lawler, J. K. Lenstra, A. Rinnooy Kan, and D. Shmoys, eds.), Wiley Interscience, New York, 1985, •A case study with emphasis on statistical techniques.

[HA80]   D. C. Hoaglin and D. F. Andrews, *The reporting of computation-based results in statistics*, The American Statistician **29** (1980), 122—126, •A statistician's view of computational experimentation.

[Hig97]   J. L. Higle, *Variance reduction and objective function evaluation in stochastic linear programs*, INFORMS Journal on Computing **10** (1997), no. 2, 236—247, •Tutorial discussion of variance reduction techniques.

[Hoo93]   J. Hooker, *Needed: An empirical science of algorithms*, Operations Research **42** (1993), no. 2, 201–212, •Hooker laments the state of algorithmic experimentation in operations research and suggests ways to make practice more rigorous.

[Hoo96]   J. N. Hooker, *Testing heuristics: We have it all wrong*, Journal of Heuristics **1** (1996), 33—42, •Proposes new approaches to algorithmic research.

[HP01]   M.G. Hall and M. E. Posner, *Generating experimental data for computational testing with machine scheduling applications*, Operations Research **49** (2001), no. 7, 854–865, Survey of good principles of data generation, together with specific critiques and illustratoins from machine scheduling.

[JBNP91]   R. H. F. Jackson, P. T. Boggs, S. G. Nash, and S. Powell, *Guidelines for reporting results of computational experiments. report of the ad hoc committee*, Mathematical Programming **49** (1991), 413–425, •A useful update to the classic article [CDM78].

[JM78]   R. H. F. Jackson and J. M. Mulvey, *A critical review of comparisons of mathemat-ical programming algorithms and software (1953-1977)*, Journal of Research of the

National Bureau of Standards **83** (1978), no. 6, 563–584, •Critique of the literature circa 1978, still relevant today.

[JM91]   D. S. Johnson and C. C. McGeoch (eds.), *Network flows and matching: First dimacs implementation challenge*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 12, American Mathematical Society, Philadelphia, 1991, Programs and test instances available from *dimacs.rutgers.edu/Challenges/"*. •A collection of experimental papers and commentary.

[Joh01]   D. S. Johnson, *A theoretician's guide to the experimental analysis of algorithms*, this volume, 2001, Also available at *www.research.att.com/˜dsj/papers/experguide.ps*. •A survey of pitfalls of experimentation and presentation, with advice on how to avoid them.

[JT96]   D. S. Johnson and M. Trick (eds.), *Graph Coloring, Vertex Cover, and Satisfiability: Second DIMACS implementation challenge*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 26, American Mathematical Society, Philadelphia, 1996, •A collection of experimental papers and commentary.

[Kel94]   W. D. Kelton, *Perspectives on simulation research and practice*, ORSA Journal on Computing **6** (1994), no. 4, 318–328.

[KG92]   J. Kleijnen and W. Van Groenendaal, *Simulation: A statistical perspective*, John Wiley & Sons, New York, 1992.

[Knu93]   D. E. Knuth, *The stanford graphbase: A platform for combinatorial computing*, ACM Press and Addison-Wesley, New York and Reading, MA, 1993, Available by anonymous ftp at `labrea.stanford.edu`. •Collection of graph algorithms and graphs with real-world structure.

[L'E90]   P. L'Ecuyer, *Random numbers for simulation*, Communications of the ACM **33** (1990), 85–97, •A standard reference on random number generators.

[L'E92]   P. L'Ecuyer, *Testing random number generators*, Proceedings of the 1992 Winter Simulation Conference (1992), 305–313.

[L'E94]   P. L'Ecuyer, *Uniform random number generation*, Annals of Operations Research **53** (1994), 77–120.

[L'E96]   P. L'Ecuyer, *Simulation of algorithms and performance analysis*, ORSA Journal on Computing **1** (1996), no. 1, 16–20, •Short commentary on [McG96b] discussing statistial analysis and random number generation.

[LED89]   *Leda research*, since 1989, •A large repository of efficient C++ implementations of algorithms and data structures.

[LR80]   B. W. Lin and R. L. Rardin, *Controlled experimental design for statistical comparison of integer programming algorithms*, Management Science **25** (1980), no. 12, 1258–1271.

[McG86]   C. C. McGeoch, *Experimental analysis of algorithms*, Ph.D. thesis, Technical Report CMU-CS-87-124, Department of Computer Science, Carnegie Mellon University, Pittsburgh, 1986, •Case studies and methodologies for algorithm experimentation.

[McG92]   C. C. McGeoch, *Analyzing algorithms by simulation: Variance reduction techniques and simulation speedups*, ACM Computing Surveys **245** (1992), no. 2, 195–212, •A tutorial on variance reduction techniques in the context of algorithm analysis, with examples.

[McG96a]   C. C. McGeoch, *Challenges in algorithm simulation*, ORSA Journal on Computing **1** (1996), no. 1, 27–28, •Brief reply to commentaries on [McG96b].

[McG96b]   C. C. McGeoch, *Toward an experimental method for algorithm simulation*, INFORMS Journal on Computing **1** (1996), no. 1, 1—15, •A survey of issues in computational experiments on algorithms.

[McG00]   C. C. McGeoch, *Experimental analysis of optimization algorithms*, Handbook of Combinatorial Optimization (P. M. Pardalos and M. G. C. Resende, eds.), Oxford University Press, 2000, A chapter surveying issues in experimentation.

[McG01a]   C. C. McGeoch, *A bibliography of algorithm experimentation*, this volume, 2001.

[McG01b]   C. C. McGeoch, *Experimental analysis of algorithms*, Handbook of Global Optimization, Volume 2: Heuristic Approaches (P. M. Pardalos and E. Romeijn, eds.), Kluwer Academic Publishers, 2001, A chapter on experimental methodology.

[MM99]   C. C. McGeoch and B. M. E. Moret, *How to present a paper on experimental work with algorithms*, SIGACT News **30** (1999), no. 4, 85–90, Also available at

`www.cs.amherst.edu/~ccm/howto.ps`. •Advice on giving talks on experimental research in algorithms.

[Mor96]   Bernard M. Moret (ed.), *Journal of Experimental Algorithmics*, *www.jea.acm.org/index.html*, since 1996, •An on-line journal published by The Association for Computing Machinery.

[Mor97]   B. M. E. Moret, *Bridging the gap between theory and practice*, September 1997, Available at *www.press.umich.edu/jep/03-01/JEA.html* •A survey of issues and opportunities in experimental algorithmics.

[Mor01]   B. M. E. Moret, *Toward a discipline of experimental algorithmics*, this volume, 2001, •General advice on experimental methodology.

[MS01]   B. M. E. Moret and H. D. Shapiro, *Algorithms and experiments: the new (and the old) methodology*, Journal of Universal Computer Science **7** (2001), no. 5, 434–446, •Historical survey of methodology.

[MSF+01]   C. C. McGeoch, P. Sanders, R. Fleischer, P. Cohen, and D. Precup, *Searching for Big-Oh in the data: Inferring asymptotic complexity from experiments*, Experimental Algorithmics, the State of The Art:, to appear as a volume of *Lecture Notes in Computer Science*, Springer Verlag, 2001, •Advice and techniques for inferring asymptotic behavior from data.

[NRLMF87]   R. E. Nance, Jr. R. L. Moose, and R. V. Foutz, *A statistical technique for comparing heuristics: An example from capacity assignment strategies in computer network design*, Communications of the ACM **30** (1987), no. 5, 430–442.

[Orl96]   J. B. Orlin, *On experimental methods for algorithm simulation.*, ORSA Journal on Computing **1** (1996), no. 1, 21–23, •Short commentary on [McG96b] focusing on performance measures.

[Paw90]   K. Pawlikowski, *Steady-state simulations of queueing processes: A survey of problems and solutions*, ACM Computing Surveys **22** (1990), no. 2, 123–170.

[PB85]   P. W. Purdom, Jr. and C. A. Brown, *The analysis of algorithms*, Holt, Reinhart and Winston, 1985, •Contains a chapter on experimental methodology.

[PM88]   S. K. Park and K. W. Miller, *Random number generators: Good ones are hard to find*, Communications of the ACM **31** (1988), no. 10, 1192–1201, •A comparison of random number generators, with advice to practioners.

[Raw98]   J. O. Rawlings, *Applied regression analysis: A research tool*, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1998, •A very readible advanced textbook.

[Rei81]   G. Reinelt, *TSPLIB–a traveling salesman problem library*, ORSA Journal on Computing **3** (1981), 376–384, Available through the mail server `netlib@research.att.com`.

[San90]   L. A. Sanchis, *On the complexity of test case generation for NP-Hard problems*, Information Processing Letters **36** (1990), 135–140.

[San93]   L. A. Sanchis, *Some complexity issues involved in the construction of test cases for NP-Hard problems*, Complexity in Numerical Optimization (Panos M. Pardalos, ed.), World Scientific, 1993.

[San95]   L. A. Sanchis, *Generating hard and diverse test sets for NP-Hard graph problems*, Discrete Applied Mathematics **58** (1995), 35–66.

[Shi96]   D. B. Shier, *On algorithm analysis*, ORSA Journal on Computing **1** (1996), no. 1, 24–26, •Short commentary on [McG96b] describing a conceptual framework for algorithm analysis.

[Ski97]   S. S. Skiena, *The Stony Brook Algorithm Repository*, *www.cs.sunysb.edu/~algorith/*, since 1997, •A repository of implementations in several programming languages.

[Ski98]   S. S. Skiena, *The algorithm design manual*, Springer-Verlag/TELOS (The Electronic Library of Science), 1998, •Much information about algorithms and applications. Great war stories.

[Spa01]   A. Marchetti Spaccamela, *Workshop on Algorithm Engineering (WAE 2001)*, *www.brics.dk/wae2001/*, Denmark, August 2001, •Fifth in an annual series of workshops. The website has links to previous meetings.

[Tod94]   M. J. Todd, *Theory and practice for interior-point methods*, ORSA Journal on Computing **6** (1994), no. 1, 28–31.

[Tuk77]   J. W. Tukey, *Exploratory data analysis*, Addison Wesley, Reading, MA, 1977, •A seminal work on non-parametric data analysis.

Department of Mathematics and Computer Science, Amherst College, Amherst, MA 01002

E-mail address: ccm@cs.amherst.edu