

Experience

January 2018 – Present **Foresite Capital** **San Francisco, CA**
Director, Data Science

- Directs Foresite's data science team, including efforts in statistical genetics, bioinformatics, and machine learning for the analysis of biomedical data

November 2016 – November 2017 **GRAIL** **Menlo Park, CA**
Staff Data Scientist

- Developed probabilistic machine learning methods for somatic variant detection and filtering
- Built quantitative methods for applications ranging from low-level assay data processing to cancer detection and classification
- Collaborated with R&D teams on data analysis, study design, and assay development
- Member of the data science team under Franz Och

June 2013 – November 2016 **Google / Alphabet** **Mountain View, CA**
Senior Statistician / Senior Scientist, Verily (January 2016 – November 2016)

- Technical lead of computational biology group
- Coordinating complex multi-omic analyses for collaborative studies
- Driving and guiding computational research with deep learning and network-based methods
- Supporting development of systems biology platform in collaboration with wet lab groups

Statistician, Google [x] / Google Life Sciences / Verily (October 2014 – December 2015)

- Drove statistical design of complex multi-omic studies, including the Baseline study
- Analyzed internally and externally generated data to support clinical and basic research, including cross-tissue expression analysis of GTEx data
- Designed and implemented core computational infrastructure with software engineering team

Statistician, Ads Metrics Team (June 2013 – October 2014)

- Worked with a team of statisticians and engineers to understand and optimize Google's search advertising (AdWords)
- Designed and analyzed experiments on AdWords systems, including long-term user learning studies
- Developed novel, scalable statistical methods, such as techniques for learning from aggregate labels

May 2008 – May 2013 **Harvard University Statistics Department** **Cambridge, MA**
Research Assistant

- Developed massively parallel MPI-based deconvolution algorithms for genome-wide analysis of chromatin structure based on high-throughput sequencing data with the O'Shea Lab (Harvard FAS Systems Biology) and the Airoidi Lab (Harvard Statistics)
- Developed a probabilistic method for absolute quantitation from LC-MS/MS data with cell biologists and MS experts
- Built Bayesian methods sparse x-ray data with scientists from the Harvard-Smithsonian Center for Astrophysics
- Developed new methods for anomaly detection & analysis in time series with Harvard Time Series Center
- Extended Bayesian approaches to network tomography

June 2008 – June 2009 **Weiss Asset Management** **Boston, MA**
Statistician (part-time)

- Developed and executed statistical analyses to drive trading strategies
- Built complex time series datasets using a range of data management tools

Sept. 2007 – June 2009 **Boston University Economics Department** **Boston, MA**
Research Assistant

- Collaborated with Laurence Kotlikoff and Stephen Ross on Social Security valuation research
- Analyzed demographic and financial data with time series techniques, hierarchical modeling, and advanced simulation methods

June 2007 – Jan. 2008
Consultant (January 2008)

UBS Investment Bank

Stamford, CT

- Developed and revised complex statistical models to improve forecasts and gain insight into bond market dynamics
- Trained quantitative strategists in the use of R for validation and revision of statistical models, including collaborative development

Summer Analyst, Fixed Income Research (June – August 2007)

- Conducted statistical analyses for US Rates and Government Bonds research desk, including VAR and VECM-based modeling, high-frequency time series analysis, regime-switching modeling, and Gaussian state space modeling
- Presented methodology and results to senior management

July – September 2006

Boston University School of Management

Boston, MA

Research Assistant

- Built database of scientific publication metadata for research on stem cell policy
- Curated historical data on private companies for study of venture capital over-investment

May 2004 – March 2007

Matté & Company Management Consulting

Greenwich, CT

Senior Research & IT Advisor (formerly Senior Research & IT Coordinator)

- Performed quantitative and qualitative research for Fortune 500-scale clients on competitors' structure and activities
- Conducted quantitative analysis and modeling for executive compensation packages
- Trained interns in research techniques and methods
- Deployed and managed IT systems to support collaborative research

Education

Harvard University

- PhD in Statistics (September 2009 – May 2013)
- AM in Statistics (September 2009 – May 2011)
- Advised by Xiao-Li Meng and Edoardo Airoldi
- Dissertation: [Distributed and Multiphase Inference in Theory and Practice: Principles, Modeling, and Computation for High-Throughput Science](#)

Boston University

- MA in Economics (September 2006 – May 2008)
- BA in Mathematics & Economics, Summa Cum Laude (September 2004 – May 2008)
- 3.97 cumulative GPA
- 4.00 GPA in mathematics & statistics coursework

Publications

- **Blocker, A. W.**, Solís, E.J., & Airoldi, E.M. (Under revision at Journal of the American Statistical Association). A model of peptide charge and modification states for accurate quantification of low concentration proteins. Draft available on request.
- Zhou, X., **Blocker, A. W.**, Airoldi, E. M., O'Shea, E. K., & Akhtar, A. (2016). A computational approach to map nucleosome positions and alternative chromatin states with base pair resolution. *eLife*, 5, e16970. <http://doi.org/10.7554/eLife.16970>
- **Blocker, A. W.**, & Airoldi, E. M. (2016). Template-based models for genome-wide analysis of next-generation sequencing data at base-pair resolution. *Journal of the American Statistical Association*, 1–68. <http://doi.org/10.1080/01621459.2016.1141095>
- Scott, S. L., **Blocker, A. W.**, Bonassi, F. V., Chipman, H. A., George, E. I., & McCulloch, R. E. (2016). Bayes and big data: the consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, 1–11. <http://doi.org/10.1080/17509653.2016.1142191>
- Wager, S., **Blocker, A.**, & Cardin, N. (2015). Weakly supervised clustering: Learning fine-grained signals from coarse labels. *The Annals of Applied Statistics*, 9(2), 801–820. <http://doi.org/10.1214/15-AOAS812>
- Goodman, A., Pepe, A., **Blocker, A. W.**, Borgman, C. L., Cranmer, K., Crosas, M., et al. (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Computational Biology*, 10(4), e1003542. <http://doi.org/10.1371/journal.pcbi.1003542>
- **Blocker, A. W.**, & Meng, X.-L. (2013). The potential and perils of preprocessing: Building new foundations. *Bernoulli*, 19(4), 1176–1211. <http://doi.org/10.3150/13-BEJSP16>
- Airoldi, E. M., & **Blocker, A. W.** (2013). Estimating Latent Processes on a Network From Indirect Measurements. *Journal of the American Statistical Association*, 108(501), 149–164. <http://doi.org/10.1080/01621459.2012.756328>

- **Blocker, A.W.** & Protopapas, P. (2012). Semi-parametric Robust Event Detection for Massive Time-Domain Databases. *Statistical Challenges in Modern Astronomy V*, Springer-Verlag, 177–189. Full version available as [arXiv:1301.3027 \[stat.AP\]](https://arxiv.org/abs/1301.3027).
- **Blocker, A. W.**, & Airoidi, E. (2011). Deconvolution of mixing time series on a graph (pp. 51–60). Presented at the Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11), Corvallis, Oregon: AUAI Press. <https://dslpitt.org/papers/11/p51-blocker.pdf>.
- **Blocker, A. W.**, Protopapas, P., & Alcock, C. R. (2009). A Bayesian Approach to the Analysis of Time Symmetry in Light Curves: Reconsidering Scorpius X-1 Occultations. *The Astrophysical Journal*, 701(2), 1742–1752. <http://doi.org/10.1088/0004-637X/701/2/1742>
- **Blocker, A.**, Kotlikoff, L., & Ross, S. (2008). *The True Cost of Social Security*. Cambridge, MA: National Bureau of Economic Research. <http://doi.org/10.3386/w14427>

Invited and Topic Contributed Talks

- “Big Data, Big Models, Big Problems: Consensus Monte Carlo Methods for Distributed Bayesian Inference.” Presented August 4 2014 at JSM 2014, Boston, MA
- “Robust Event Detection for Massive Time-Series Surveys.” Presented July 30 2012 at JSM 2012, San Diego, CA. Slides available at http://www.awblocker.com/Blocker_JSM2012.pdf
- “Preprocessing, Multiphase Inference, and Massive Data in Theory and Practice.” Presented July 13 2012 at MMDS, Stanford University, Stanford, CA. Slides available from MMDS at <http://www.stanford.edu/group/mmds/slides2012/s-blocker.pdf>
- “Deconvolution of Mixing Time Series on a Graph.” Presented April 16 2011 at NESS 2011, University of Connecticut, Storrs, CT
- “Semi-parametric Robust Event Detection for Massive Time-Series Datasets.” Presented August 25, 2010 at the Workshop on Computational Astrostatistics, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA (<http://hea-www.harvard.edu/AstroStat/CAS2010/>). Slides available at http://www.awblocker.com/Blocker_CAS2010_20100825.pdf
- “Discussion: The Promise & Peril of Synthetic & Integrated Data.” Presented June 21, 2010 at ISCA Applied Statistics Symposium, Indianapolis, IN. Slides available at http://www.awblocker.com/discussion_AWB_20100621.pdf
- “Doing Right By Massive Data: Using Probability Modeling To Advance The Analysis Of Huge Astronomical Datasets.” Presented April 17, 2010 at NESS 2010, Harvard University, Cambridge, MA. Slides available at http://www.awblocker.com/discussion_AWB_20100621.pdf

Technical Skills

- Extensive experience with biological data analysis, including high-throughput sequencing, MS/MS proteomics, epigenetic assays, and clinical data
- Advanced proficiency in Go, Python, R, and \LaTeX
- Proficient with C/C++ (with BLAS/LAPACK, MPI, and Eigen) and Java
- Experience with Google-scale distributed computing via MapReduce, Flume, and related systems
- Experience developing scientific applications in Python, R, C/C++, and Matlab including:
 - Parallel deconvolution algorithms for the analysis of high-throughput genetic sequence data for chromatin structure research
 - Bayesian simulation algorithms for astrophysical data in R and Python (e.g. bayesstack)
 - Event detection algorithms for massive time series datasets in C with BLAS/LAPACK with application to astronomical surveys
 - Particle filtering / MCMC methods for network tomography in R and C, including methods for efficient simulation from distributions on convex polytopes
 - EM algorithms for the estimation of linear dynamical systems with covariates in Matlab (available on website)
- Experience with estimation, hierarchical modeling, simulation, optimization, and high-performance (including parallel) computing in Python, R, and C/C++
- Thorough knowledge of Linux, including the development of EC2- and Google Cloud Platform-based systems for distributed scientific computing
- Experience with database development using PostgreSQL, SQLite, BigQuery, and Google internal systems

Awards & Achievements

Paper Awards

- Arthur P. Dempster Award (inaugural) for “Preprocessing and Multiphase Inference in Theory and Practice”; awarded May 2012

- NESS Best Paper Award for “Template-based methods for analyzing chromatin structure dynamics genome-wide” with Edo Airoidi; awarded at NESS 2012 in April 2012
- IBM Thomas J. Watson Research Center Student Research Award for “Deconvolution of mixing time series on a graph” with Edo Airoidi; awarded at NESS 2011 in April 2011
- NESS Student Paper Award (Sponsored by Microsoft and Google) for “A Bayesian Approach to the Analysis of Time Symmetry in Light Curves: Reconsidering Scorpius X-1 Occultations” with Pavlos Protopapas; awarded April 2010

Teaching Awards

- Harvard University Certificate of Distinction in Teaching for Statistics 221, Fall 2010; awarded April 2011

Academic Honors

- Phi Beta Kappa Initiate, May 2008
- Boston University College Prize for Excellence in Economics, May 2008
- Phi Beta Kappa Award recipient, December 2006
- BU Distinguished Sophomore in College of Arts & Sciences

Fellowships

- Pierce Fellowship Recipient, Harvard University Graduate School of Arts & Sciences, September 2009
- BU University Scholarship recipient (merit-based)

Teaching Experience

Sept. 2008 – May 2013 **Harvard University Statistics Department** **Cambridge, MA**

Head Teaching Fellow for Statistics 104 (Fall 2012)

- Responsible for sectioning and other administrative matters for class of 484 students with a team of 16 teaching fellows
- Taught weekly discussion section to reinforce lecture material

Teaching Fellow for Statistics 211 (Spring 2012)

- Taught 2 sections for “Statistical Inference”, a core PhD curriculum course

Teaching Fellow for Statistics 244 (Fall 2011)

- Taught 1 section for “Linear and Generalized Linear Models”, the first offering of this course at Harvard

Teaching Fellow for Statistics 111 (Spring 2011)

- Taught 2 sections for “Introduction to Theoretical Statistics”
- Wrote all problem sets and led exam writing

Teaching Fellow for Statistics 221 (Fall 2010)

- Taught 2 sections for “Statistical Computing & Learning”
- Covered details of implementing high-dimensional numerical optimization, MCMC, EM, and LDA
- Guided students through replicating the computation from one JASA/JMLR paper per biweekly problem set

Head Teaching Fellow for Statistics 104 (Spring 2009)

- Responsible for sectioning and other administrative matters for class of 167 students with a team of 8 teaching fellows
- Developed section notes for other teaching assistants to guide weekly sections
- Taught discussion section (17 students total) to reinforce lecture material

Teaching Assistant for Statistics 104 (Fall 2008)

- Taught 2 discussion sections (34 students total) per week to reinforce lecture material
- Held office hours to address individual questions and concerns
- Graded homework and exams from assigned students

Organizations

- Phi Beta Kappa
- American Statistical Association
- Institute for Mathematical Statistics
- Bernoulli Society

Background

- US Citizen

References available upon request