

LAUREL LIANE STELL

Biostatistician, Department of Biomedical Data Science, Stanford University

EDUCATION

M.S. Statistics, Stanford University, Stanford, CA, April 2014

Ph.D. Applied Mathematics, Cornell University, Ithaca, NY, January 1993

M.S. Applied Mathematics, Cornell University, Ithaca, NY, January 1989

B.S. Mathematics, New Mexico Institute of Mining and Technology, Socorro, NM, December 1985

SUMMARY OF COMPUTER SKILLS

Proficient with R (including R markdown, Bioconductor tools, and multicore functionality), Perl, and UNIX shell scripts and utilities. Extensive UNIX/Linux system administration. Prior extensive experience with MATLAB, Java, C++, C, and FORTRAN.

PROFESSIONAL EXPERIENCE

Department of Biomedical Data Science, Stanford University, Stanford, CA 94305

Engineering Research Associate, September 2013–present

Summary: Biostatistician with expertise in statistical modeling and inference, variable selection, and multiple hypothesis testing. Data types include derived genotypes, DNA sequencing including whole exome, mRNA-seq, gene expression microarray, methylation information, novel high-resolution nanoimmunoassay of tumor tissue, and pollution sensor data.

- **Estimating latent confounders to improve differential expression testing of mRNA-seq data.** Currently the sole statistician on a project using bulk RNA-seq data from human samples in four disease states towards the goal of developing biomarkers for Type 1 diabetes (T1D). PCA plots and distributions of the differential expression (DE) test statistics indicated the existence of latent factors. When surrogate variable analysis (SVA) failed to give satisfactory results, reviewed literature on methods to estimate the number of confounders as well as estimating the confounders themselves. The newer R package *cate* gives better results than SVA when the correlation between the factors and the variable being tested for DE is sufficiently strong. Compared results for the two estimates of hidden factors, using *limma*, *DESeq2* and *cate* to test for DE. Using the confounder estimates from *cate* gave much more plausible PCA plots, number of discoveries, and top-ranked genes for DE between disease states. To provide further insight, also tested for DE between sexes and found *cate* to be best at prioritizing genes on X and Y chromosomes over autosomal genes.
- **Currently consultant for medical research projects involving laboratory experiments as well as both retrospective medical record reviews and prospective human studies.** Clients range from graduate students to postgraduate trainees to physicians. Level of involvement ranges from one-hour consultations to writing statistical methods for proposals to co-author on manuscripts. Statistical tools have included ordinary and logistic regression, mixed-effects models, lasso penalization with cross-validation, multivariate analysis such as PCA and hierarchical clustering, data smoothing, simulations, and novel visualization approaches.
- **False discovery control when testing for genotype-phenotype associations with whole exome sequencing and 60 metabolomics measures.** Part of a team from five universities that tested associations between about 60 metabolomics measures and 500,000 variants from whole exome sequencing of 20,000 Finns. Primary responsibility was to develop a hypothesis testing procedure that accounts for not only many variants but also many phenotypes to control false discoveries. Adapted a hierarchical approach, originally introduced by Benjamini and Bogomolov, that uses p-values from testing separately association of each variant with each phenotype. Engaged extensively with the rest of the team on other issues, particularly data visualization to guide research decisions and for the manuscript, currently resubmitted.
- **Models for error in novel nanoimmunoassay and for intra- and inter-tumoral variability.** Part of a team led by two physicians to investigate biomarkers for renal cell carcinoma using a novel high-resolution nanoimmunoassay. Worked closely with laboratory associate to assess potential metrics and address issues such as detectability and reproducibility. Developed model for intra- and inter-tumoral heterogeneity and performed simulations to demonstrate that data were consistent with model. Provided statistical methods and figures for manuscript, and also provided methods and simulation-based power calculations for proposals for follow-up funding.

- **Developed a method for Bayesian variable selection using hierarchical priors with application to genotype-phenotype associations.** Used linear regression model with a prior that combines information across responses and one that combines information across covariates, as well as a standard spike and slab prior for comparison. Tested with simulations as well as discovering associations in real genotype and phenotype data. Lead author on manuscript and sole developer of the R package `ptycho` published on Comprehensive R Archive Network (CRAN).
- **Modeled relationships between asthma, DNA methylation, pollution exposure, seasons and family effects.** Part of a team led by a physician to investigate relationship between asthma, DNA methylation, and exposure to air pollutants over various time intervals. Worked with the laboratory associate to design the data tables and resolve irregularities in the data. Identified techniques to handle noise and missing values in EPA pollution data. Performed linear regression to analyze associations, checking whether the methylation might be better explained by family (modeled as random effect) or season.

**NASA Ames Research Center, Moffett Field, CA
Aerospace Engineer, September 2007–August 2013**

Summary: Conducted research in air traffic management.

- Responsible for developing aircraft trajectory predictor to enable fuel-efficient descents for arrivals in congested airspace. Worked with FAA controllers and airline pilots to develop protocols for an exercise to collect operational data. Cleaned and analyzed the data, then compared existing kinetic predictor to a linear regression model with variables selected by cross-validation.

**Metron Aviation, Inc., Herndon, VA
Senior Analyst, May 2003–August 2007**

Summary: Used mathematical modeling, primarily empirical, to address problems related to air traffic management.

- Worked on the NASA-funded Airport Configuration Predictor SBIR as the primary analyst in Phase 1 and as the Principal Investigator in Phase 2. Assessed the accuracy of possible sources of runway configuration data, including algorithmic determination of configuration from surveillance data. Surveyed factors affecting runway usage, including a statistical analysis with a logistic regression model. Developed an optimization model for runway usage and devised heuristic computation techniques to reduce solution times.
- Developed a multiple hypothesis Kalman filter to use ASDE-X airport surface surveillance data to estimate aircraft position and speed on a taxiway or runway and to extract events such as takeoffs, landings, and entering or exiting the airport movement area. Designed, implemented, and tested a prototype in MATLAB. Provided support to the software engineer who translated it to Java for operational testing by the FAA.

**Metron, Inc., Reston, VA
Analyst, October 1999–May 2003**

Summary: Worked on every stage of code development—from mathematical modeling through algorithm development to implementation and testing—for scientific applications for the US military.

- Responsible for all facets of the development of the mathematical algorithms in a naval exercise reconstruction application for operational use by the Surface Warfare Development Group. Estimated vessel state by applying a Kalman smoother to position and velocity data, adapting the filter to maintain numerical stability without requiring an initial state estimate.
- Worked on a project to develop a vessel tracking algorithm for US Navy submarines using a genetic implementation of a nonlinear Bayesian tracker to fuse the linear towed array and spherical array sonar reports. Simulated the motion of the towed array using a finite difference scheme, adapting the spatial discretization to maintain stability as the speed changed.

**Georgia Institute of Technology, Atlanta, GA
Systems Support Specialist III, Office of Information Technology,
November 1997–September 1999**

Summary: Sole system administrator for two multiple-processor SGI Origin 2000s and an Origin 200 available campus-wide for scientific computation.

- Responsible for software installation, testing, and maintenance, including high performance utilities.
- Analyzed effects such as architecture, compiler optimization, and parallel programming model on performance. Read assembler code to identify compiler optimizations.

Systems Support Specialist II, School of Mathematics, November 1996–November 1997

Summary: System administrator for a network of about 75 Sun Solaris computers.

Hazlehurst and Associates, Inc., Atlanta, GA

Consultant, November 1995–November 1996

Summary: Provided UNIX expertise for an OS/2 shop investigating conversion to an IBM AIX server.

Georgia Institute of Technology, Atlanta, GA

Temporary Assistant Professor, School of Mathematics, 1992–1995

PUBLICATIONS AND CONFERENCES

A. Locke, K. Steinberg, C. Chiang, S. Service, A. Havulinna, L. Stell, et al (resubmitted) “Exome sequencing identifies high impact trait-associated alleles enriched in Finns”, <http://biorxiv.org/cgi/content/short/464255v1>

C. Hoerner, R. Massoudi, T. Metzner, L. Stell, et al (in press) “Multiregion quantification of extracellular signal-regulated kinase activity in renal cell carcinoma”, *European Urology Oncology*, <https://doi.org/10.1016/j.euo.2018.09.011>

M. Prunicki, L. Stell, D. Dinakarbandian, et al (2018) “Exposure to NO₂, CO, and PM_{2.5} is linked to regional DNA methylation differences in asthma”, *Clinical Epigenetics* **10**:2–12.

L. Stell and C. Sabatti (2016) “Genetic variant selection: learning across traits and sites”, *Genetics* **202**:439–455.

L. Stell, J. Bronsvort, G. McDonald (2013) “Regression analysis of top of descent location for idle-thrust descents”, 10th USA/Europe Air Traffic Management R&D Seminar, Chicago, IL.

L. Stell (2011) “Flight Management System execution of idle-thrust descents in operations”, 30th Digital Avionics Systems Conference (IEEE DASC 11), Seattle, Washington.

L. Stell (2011) “Prediction of top of descent location for idle-thrust descents”, 9th USA/Europe Air Traffic Management R&D Seminar, Berlin, Germany.

L. Stell (2010) “Analysis of Flight Management System predictions of idle-thrust descents”, 29th Digital Avionics Systems Conference (IEEE DASC 10), Salt Lake City, Utah.

L. Stell (2010) “Predictability of top of descent location for operational idle-thrust descents”, AIAA 10th Aviation, Technology, Integration, and Operations (ATIO) Conference, Fort Worth, Texas.

R. Copenbarger, G. Dyer, M. Hayashi, R. Lanier, L. Stell, D. Sweet (2010) “Development and testing of automation for efficient arrivals in constrained airspace”, 27th International Congress of the Aeronautical Sciences (ICAS 2010), Nice, France.

L. Stell (2009) “Flight Management System prediction and execution of idle-thrust descents”, 28th Digital Avionics Systems Conference (IEEE DASC 09), Orlando, Florida.

L. Stell, T. Carniol, S. Lozito, V. Sud (2005) “Flight state estimation from surface surveillance”, 5th NASA Integrated Communications, Navigation and Surveillance (ICNS) Conference and Workshop, Fairfax, Virginia.

S. Atkins, Y. Jung, C. Brinton, L. Stell, T. Carniol, S. Rogowski (2004) “Surface management system field trial results”, 4th Annual AIAA Aviation, Technology, Integration, and Operations (ATIO) Forum.

L.L. Stell and S.F. Shen (1995) “A fixed domain method for injection governed by the Stokes equations”, *SIAM J. Scientific Computing* **16**:798–818.