

Daniel V. Samarov
870 Wildwood Rd
NE Atlanta, GA
30324 dan@dsbox.co
732-322-2565

Education

- **PhD in Statistics** (2009), Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC. Dissertation: The Analysis and Advanced Extensions of Canonical Correlation Analysis Advisors: J. Steven Marron, Yufeng Liu and Alexander Tropsha.
- **BA in Mathematics and Statistics** (2002), Rutgers College, New Brunswick, NJ.
- **BA in Economics** (2002), Rutgers College, New Brunswick, NJ.

Professional Experience

- **Cofounder, Partner DS**
Box (2018-Present).
- **Mathematical Statistician** Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD (2009-present).
- **Cofounder, Chief Data Scientist** RxAnalytics (2011-2014).
- **Data science consultant** Private (2014-present).
- **Graduate Research Assistant** Becton Dickinson, Research Triangle Park, NC (2006-2009).
- **Consultant** Becton Dickinson, Research Triangle Park, NC (2006-2009).
- **Teaching Assistant** University of North Carolina at Chapel Hill, Chapel Hill, NC (2004-2006).
- **Underwriter** Chubb Group of Insurance Companies, Whitehouse Station, NJ (2002-2004).

Publications, submissions and works in progress

- Burch CL, Guyader S, **Samarov DV** and Shen H. (2007) “Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus phi 6.” *Genetics* 176: 467-76.
- **Samarov DV** (2007), “The Analysis and advanced extensions of canonical correlation analysis”, dissertation, The University of North Carolina at Chapel Hill.
- Howard-Reed C, **Samarov DV**, Leber D, Hodgson A, Mason S, and Won D. (2011) “Developing a Diffusive Reference Material for VOC Emissions Testing: Pilot Interlaboratory Study.” *Building and Environment* 46: 1504-1511.
- Rukhin A and **Samarov DV**. (2011) “Limit of Detection Determination for Censored Samples,” *Chemometrics and Intelligent Laboratory Systems* 105: 188-194.
- Mauchant D, Rice KD, Riley MA, Leber D, **Samarov DV** and Forster AL (2011) “Analysis of Three Different Regression Models to Estimate the Ballistic Performance of New and Environmentally Conditioned Body Armor” NISTIR 7760.
- Reed CH, Liu Z, Benning J, Cox S, **Samarov DV**, Leber D, Hodgson AT, Mason S, Won D and Little JC. (2011) “Diffusion-controlled reference material for volatile organic compound emissions testing: Pilot inter-laboratory study” *Building and Environment* 46: 1504-1511.
- Litorja M, Allen D, Hwang J, **Samarov DV** and Clarke M. (2011) “Characterization of hyperspectral imaging and analysis via microarray printing of dyes,” *SPIE Photonics West*.
- Reed C, **Samarov DV**, Leber D, Liu Z, and Little J. (2011) “Assessing the shelf-life of a prototype reference material for product emissions testing,” *Indoor Air*.
- Douglass KO, Maxwell SE, Plusquellic DF, Hodges JT, van Zee RD, **Samarov DV**, Whetstone JR. (2011) “Construction of a High Power OPO Laser system for Differential Absorption LIDAR”, *SPIE Lidar Remote Sensing for Environmental Monitoring* 8159: 81590D1-9.
- **Samarov DV**, Marron JS, Liu Y, Grulke C, Tropsha A. (2011) “Local Kernel Canonical Correlation Analysis with Application to Virtual Drug Screening,” *Annals of Applied Statistics* 5: 2169-2196.
- **Samarov DV**, Hwang J, Lee JY and Clarke M. (2012) “Validating the LASSO algorithm for unmixing spectral signatures with application to multicolor phantoms”, *Conference Proceedings SPIE Photonics West*.
- Clarke ML, Lee JY, **Samarov DV**, Allen DW, Litorja M and Hwang J. (2012), “Designing Microarray Phantoms for Hyperspectral Imaging Validation,” *Biomed Opt Express* 3:12921-9.
- **Samarov DV**, Hwang J, Lee JY and Clarke M. (2012) “Algorithm Validation using Multicolor Phantoms,” *Biomedical Optics Express* 3:1300-11.
- Zook JM, **Samarov DV**, McDaniel J, Sen SK and Salit M. (2012) “Synthetic spike-in standards

improve run-specific systematic error analysis for DNA and RNA sequencing.” PLoS ONE 7:e41356.

- Levine ZH, Gerrits T, Migdall AL, **Samarov DV**, Calkins B, Lita AE, Nam SW. (2012) “An algorithm for finding clusters with a known distribution and its application to photon-number resolution using a superconducting transition-edge sensor.” J Opt Soc Am B 29: 2066-2073.
- Hechenbleikner EM, Makary MA, **Samarov DV**, Bennett JL, Gearhart, SL, Efron JE and Wick EC. (2013). “Hospital Readmission by Method of Data Collection,” J Am Coll Surg, in press.
- **Samarov DV**, Wehner E, Schwartz R, Zuzak K, Livingston E. (2013) “Hyperspectral image segmentation of the common bile duct”, SPIE Photonics West.
- Reed CH, Liu Z, Cox S, Leber D, **Samarov DV**, Little J. (2013) “Diffusion-controlled Reference Material for VOC Emissions Testing: International Inter-laboratory Study”, Journal of the Air and Waste Management Association.
- Hechenbleikner E, Miller J, Leung C, **Samarov DV**, Gearhart S, Efron J, MD, Fang S, Wick E. (2013) “Care Coordination Interventions to Prevent Readmission in Surgical Patients: What is the Role?” American J. Med. Qual.
- Katchy A, Pinto C, Jonsson P, Vu TN, Pandelova M, Riu A, Schramm KW, **Samarov DV**, Gustafsson JA, Bondesson M, Williams C, (2013) “Co-exposure to Phytoestrogens and Bisphenol a Mimics Estrogenic Effects in a Cooperative Manner”, Toxicological Sciences.
- **Samarov DV**. (2014) “The fast RODEO for Local Polynomial Regression”, Journal of Computational and Graphical Statistics.
- **Samarov DV**, Litorja M, Hwang J. (2014) “The Spatial LASSO with Applications to Unmixing Hyperspectral Biomedical Images”, Technometrics.
- Cambaliza MO, Shepson PB, Caulton D, Stirr B, **Samarov DV**, Gurney K, Turnbull J, Davis KJ, Possolo A, Moser B, Lauvaux T, Mays K, Sweeney C, Whetstone J, Karion A, Huang J, Razlivanov I, Miles NL, Richardson SJ, Hendricks A. (2014) “Assessment of uncertainties of an aircraft-based mass-balance approach for quantifying urban greenhouse gas emissions”, Atmospheric Chemistry and Physics.
- **Samarov DV**, Allen D, Litorja D, Hwang J (2016) “A coordinate descent based approach to solving the sparse Group Elastic Net”, Technometrics.
- **Samarov DV** (2015), “The Sparse Spatial Group LASSO”, Journal of Computational and Graphical Statistics (to be submitted).
- **Samarov DV**, Lee YL (2015), “Characterizing spectral signals and distributions in CARS data” (in progress).
- **Samarov DV**, Over P, Voorhees E, Soboroff I (2015), “Using replicates to improve the evaluation of information retrieval systems”, Transaction on Information Systems (second round of review).
- Olson, DO, Zook JM, Jackson, SA, **Samarov DV**, Salit, ML (2015), “PEPR: Pipeline for

Evaluating Prokaryotic References”, PeerJ.

- Litorja M, **Samarov DV** (2015), “Use of Image Segmentation Algorithms to Test Model Spectral Distribution for Surgical Lighting”. Bioimaging Conference.
- Kusne, G, **Samarov DV** (2016), “High-Throughput Determination of Structural Phase Diagram and Constituent Phases using Prior Knowledge with the Sparse Spatial Group LASSO”, (in progress).
- **Samarov DV**, Kusne, G (2016), “Group sparse matrix factorization based approaches to cluster number identification”, (in progress)
- **Samarov DV**, Over P, Voorhees E, Soboroff I (2016), “Using replicates to improve the evaluation of information retrieval systems”, 2016 Special Interest Group on Information Retrieval (SIGIR).
- Ak R, **Samarov DV**, Bhinge R (2016), “A Comparative Analysis of Regressions Models for Energy Prediction of Milling Machine Tools”, American Society of Mechanical Engineers.
- Hwang, J, Cheney P, McClatchy D, **Samarov DV**, Hyun-Jin K, Kanick S, Pogue B (2016), “Label-free Hyperspectral Dark-Field Microscopy for Quantitative Scattering Imaging of Tissue-Mimicking Phantoms”, SPIE BiOS 2016 conference.
- **Samarov DV**, Leber D, Guthrie, W (2017) “A Scalable Recommendation System for Equipment Calibration using Factorization Machines”, (in progress).
- Sisco ER, Najarro MN, **Samarov DV**, Lawrence JA, “Quantifying the stability of trace explosives under different environmental conditions using electrospray ionization mass spectrometry”, Talanta (2017).
- Kotevska O, Kusne G, **Samarov DV**, “Dynamic Network Model for Smart City Data-loss Resilience Case Study: City-to-City Network for Montgomery County Crime Analytics”, IEEE Access (2017).
- Tonner, P, Kusne G, **Samarov DV**, “Designing over uncertain outcomes with stochastic sampling Bayesian Optimization”, NeurIPS 2019 (under review).

Report

s

- **Samarov DV**, Liu Y, Marron JS (2010). “High Dimension Low Sample Size Asymptotics of Canonical Correlation Analysis”. Technical report.
- **Samarov DV**, Urbas A, Choquette S. (2012) “Raman Spectroscopy Standard: Report of statistical analysis for Standard Reference Material 2246”.
- **Samarov DV**, Fagan J, Walker AH. (2012) “Determination of Carbon Nanotube Homogeneity in Candidate Single-Wall Carbon Nanotubes (Dispersed, Three-Length Resolved Populations): Reference Material 8281”.
- **Samarov DV**, J. Zook (2014), “Sequencing error and multiple comparison corrected read depth sample size estimation”.

- **Samarov DV**, Urbas A, Choquette S. (2015) “Stability Assessment Report of Analysis for Standard Reference Material 2244”.

Seminars, Presentations,
Courses

- Statistics and Operations Research Colloquia, UNC, Chapel Hill, NC (2008): “Analysis and Advanced Extensions of Canonical Correlation Analysis”.
- Bioinformatics and Computational Biology Reception, UNC, Chapel Hill, NC (2008): Presented poster, “Analysis and Advanced Extensions of Canonical Correlation Analysis”.
- National Cancer Institute, Biometric Research Branch (2009): “Analysis and Advanced Extensions of Canonical Correlation Analysis”.
- Statistical Engineering Division (SED) Seminar, NIST, Gaithersburg, MD (2009): “Analysis and Advanced Extensions of Canonical Correlation Analysis”.
- American Society for Testing and Materials, St. Louis, MI (2010): Statistical Analysis of pilot Volatile Organic Compound study.
- UseR Conference, Gaithersburg, MD (2010): “Tutorial on StatET plug-in for the Eclipse”.
- UseR Conference, Gaithersburg, MD (2010): “Generalized Significance in Scale Space: The GS3 Package”.
- American Geophysical Union, San Francisco, CA (2010): Presented poster, “Fast, Locally Adaptive, Interactive Retrieval (FLAIR) Algorithm for the Analysis of DIAL Measurements”.
- Optical Medical Imaging Workshop, NIST, Gaithersburg, MD (2011): “Aspects and Challenges of Hyperspectral Images”.
- Optical Medical Imaging Workshop, NIST, Gaithersburg, MD (2011): “Development and Assessment of Algorithms for Hyperspectral Image Analysis”.
- Reference Material Development for Product Emissions Testing Workshop, NIST, Gaithersburg, MD (2011): “Analysis of two studies of Volatile Organic Compound emissions”.
- SED Seminar, NIST, Gaithersburg, MD (2011), “The spatial LASSO with applications to hyperspectral images”.
- Greenhouse Gas Initiative, NIST, Gaithersburg, MD (2011): “The application of the Fast RODEO to the Purdue Influx Experiment”.
- National Research Council Panel Review, NIST, Gaithersburg, MD (2011): The Greenhouse Gas Initiative and Optical Medical Imaging Project: Summary of statistical challenges and current approaches.
- Brinell/Rockwell hardness group, NIST, Gaithersburg, MD (2011): “Determination of Indentation Depth Using Local Polynomial Regression”.
- Office of Law Enforcement Standards, Gaithersburg, MD (2011): Introduction to R and StatET,

jointly with Possolo A and Leber D.

- SED Seminar, NIST, Gaithersburg, MD (2012): “The Spatial Lasso with Applications to Unmixing Hyperspectral Images”.
- SPIE Photonics West, San Francisco, CA (2012): “Validating the Lasso Algorithm by Unmixing Spectral Signatures in Multicolor Phantoms”.
- Briefing on Big Data presented to members of the Office of the Director (Gallagher P, May W, Romine C) and to the NIST big data working group, NIST, Gaithersburg, MD (2012).
- Information Technology Laboratory (ITL), Innovations in Measurement Science task force, NIST, Gaithersburg, MD (2012): “Measurement Science of Massive Data”.
- Joint Statistical Meeting, San Diego, CA (2012): “The Spatial Lasso with Applications to Unmixing Hyperspectral Images”.
- ITL Science Day, NIST, Gaithersburg, MD (2012): Presented poster, “FLAIR Algorithm for the Analysis of DIAL Measurements”.
- SPIE Photonics West, San Francisco, CA (2013): “Hyperspectral image segmentation of the common bile duct”, paper presented by K. Zuzak.
- Joint Research Conference (2014): “Sequencing Error Adjusted and Multiple Comparison Corrected Read Depth Estimation” (to be presented).
- ITL Town Hall Meeting (2013): “Things you need to know about: ‘NIST’s response to the OSTP Data Access Memorandum”.
- SED Seminar, NIST, Gaithersburg, MD (2014): “Estimating the number of unique spectral signatures in a Hyperspectral Images”.
- SED Seminar, NIST, Gaithersburg, MD (2014): “Sequencing Error Adjusted and Multiple Comparison Corrected Read Depth Estimation”
- Optical Medical Imaging Workshop, NIST, Gaithersburg, MD (2014): “Estimating the number of unique spectral signatures in a Hyperspectral Images.”
- NIST Greenhouse Gas Program Review, NIST, Gaithersburg, MD (2014): “Adaptive smoothing methods for GHG analysis”.
- TREC Conference, NIST, Gaithersburg, MD (2014): “Assessment of system performance in information retrieval” (talk and poster).
- Virginia Commonwealth University, Richmond, VA (2014): “Sparse approaches to unmixing hyperspectral images”.
- Virginia Commonwealth University, Richmond, VA (2014): “Opportunities in statistics and data science”.
- ITL Innovations in Measurement Science, NIST, Gaithersburg, MD (2014): “Got Big Data – Got Many Dimension – What now?”
- World Metrology Day, NIST, Gaithersburg, MD (2015): “Quantitative Optical Medical Imaging”

(poster). Hwang J, Allen D, Litorja M, Lemaillot P, **Samarov DV**.

- NRC Revue, NIST, Gaithersburg, MD (2015): “Adaptive smoothing methods for Greenhouse Gas Analysis”.
- Green House Program Revue, NIST, Gaithersburg, MD (2015): “Distributed Computing for Big Data Greenhouse Gas Applications”.
- Telogical Systems (2015): “Personalized Channel Recommendation Engine”.
- Litorja M, Samarov DV (2015), “Use of Image Segmentation Algorithm to Test Model Spectral Distribution for Surgical Lighting”. Bioimaging Conference (poster).
- Samarov DV, Rapp C (2016), “Matching algorithm for A360”. 2016 Abundance 360 Summit.
- Ak R, Samarov DV, Bhinge R, “A comparative analysis of regression models for energy prediction of milling machine tools”. Verification and Validation Symposium, Las Vegas, NV.
- Kusne G, Belianinov A, Samarov DV, Takeuchi, I (2016 and 2017), “Machine Learning for Material Research: Bootcamp and Workshop 2016 (2017)”, University of Maryland, College Park (<https://www.nanocenter.umd.edu/events/mlmr/>).
- Samarov DV, Leber D (2017), “Predicting equipment calibration times with Factorization Machines”, NAVAIR, Pax River, MD.
- Samarov DV, Sisco E, Najarro MN (2017), “Quantifying the stability of trace explosives under different environmental conditions using electrospray ionization mass spectrometry”, NIST Forensics Program Review, Gaithersburg, MD.
- Kusne G, Belianinov A, Samarov DV, I (2017), “Machine Learning for Material Research”, MRS Fall Meeting 2017, Boston, MA (<http://www.mrs.org/fall2017>).
- Samarov, DV (2017), “Machine Learning and it’s Application to Digital Advertising”, ADEXT, Mexico City, Mexico.
- Kusne G, Samarov DV (2018), “Machine Learning for Material Research”, Carnegie Mellon University, Pittsburgh, PA (<https://events.mcs.cmu.edu/mlse/machine-learning-for-materials-research-bootcamp/>).
- Samarov, DV, Leber, D (2018), “Predicting turn around time for naval equipment calibration using XGBOOST”. NAVAIR.
- Kusne G, Belianinov A, Samarov DV, Takeuchi, I (2018), “Machine Learning for Material Research: Bootcamp and Workshop 2018”, University of Maryland, College Park (<https://www.nanocenter.umd.edu/events/mlmr/>).
- Samarov DV (2018), “Machine Learning at ADEXT”, Technical Due Diligence
- Kusne G, Samarov DV (2018), “Machine Learning for Chemists: NSF Bootcamp and Workshop 2018”, Airlie Resort.

National Institute of Standards and Technology (NIST) (2009 -

Current)

- Development, publication and presentation of novel statistical methodology addressing NIST's mission for supporting innovation and development of standards (see publications and presentations above for details).
- Provided general consultation, mentorship and guidance on statistical methods to NIST staff.
- Assisted in the development and assessment of standard reference materials.
- Established working collaborations with staff across disciplines to include IT, biological, physical, chemical and material sciences.
- Served on committees and panels as a representative of NIST's Statistical Engineering Division (SED).
- Served on conference committees and organized visits of invited guest speakers.
- Graduate and post-doctoral advisor/mentor.
- NRC Advisor (2018 - present)

RxAnalytics – Science Behind Sweat (2011-2014)

- A subscription based (annual fee) web application (old site: <http://sciencebehindsweat.com>, beta: <http://beta.sciencebehindsweat.com>) for CrossFit (a fitness/training program, <http://crossfit.com>) focused on providing individual athletes and trainers/coaches a user friendly way to log, track and analyze their data.
- Developed and implemented statistical/machine learning methods and visualizations for summarizing and assessing individual and gym level progress, and providing automated feedback/recommendations based on identified problem areas/user goals.
- Platform built on Ruby, Rails, R, MongoDB, jQuery (contributed on both front and back end development).
- Involved in outreach and collaboration with experts in the fields of personal training and physical therapy.
- Member of 2012 Triangle Startup Factory class (<http://trianglestartupfactory.com>) (\$50k funding, \$20k post-incubator funding).

Consultin

g

- **PHD Ventures** (Dec 2014 - Feb 2015, Sep 2015 - Feb 2016):
 - o Developed a machine learning based recommendation systems for Adundance 360 attendees (<http://www.abundance360summit.com/>) highlighting whom they should meet/network with at the conference (<https://www.linkedin.com/pulse/how-mine-your-data-peter-diamandis>).

- Identify most similar attendees
 - Constructing groupings (clusters) of ~7-8 attendees and identifying unique features of each cluster.
 - Used feedback from previous year to build a semi-supervised model for the 2016 summit.
 - Presented algorithm at the 2016 Abundance 360 Summit
- Created UI using Shiny (<http://shinyapps.io>).

- **Telological Systems /Adacus** (Mar 2015 - Present):

- o Developed algorithms for (implicit) personalized ad/product recommendations.

Methodology was based on online logistic regression (sparse follow the leader regularization, factorization machines with delayed conversion windows).

- A/B tests showed consistent 20-50%+ increase in sales conversions in production.
- Provided guidance/education on methods, presented results, reviewed pitch decks to prospective clients and investors.
- Addressed general questions in the ad recommendation space.

Channel recommender system for customer television viewing habits.

- An ML model based on matrix factorization was used to predict new content (channels) customers may be interested in viewing.

- Reviewed/consulted on client presentations and pitches.
- o Developed a Bayesian A/B testing, sequential testing and sample size estimation platform

(currently used in production).

- o Research into uplift models/decision trees for audience insights.

- o Conversion attribution using Shapley Value Logistic Regression

- o General guidance/education on ML related questions, data science and analytics.

- o ML and analytics components of this project played a key role in the acquisition of Adacus

(<https://adexchanger.com/the-creative/thunder-acquires-adacus-bringing-targeting-measurement-automation-platform/>) by Thunder (<http://makethunder.com>).

- **Experfy** (Nov 2014 - Present)

- o <https://www.experfy.com/blog/dive-into-your-data-some-insight-about-an-efficient-variable-selection-process>.
- o Worked as part of hand selected team to map out potential analytics capabilities for a

major telecom company.

- o Provided general guidance on machine learning and data science projects.

- o Currently developing an online curriculum for supervised learning focused on

methods

- [Starling Trust](#) (Oct 2015 - Dec 2015)
 - o Developed and implemented a graphical model for analyzing trust between groups of individuals.
 - o Data included structured and unstructured data types.
 - o Provided visualization of results and worked with development team to implement graph/network analytics and visualization into their production environment
- [Rasa](#) (April 2016 - June 2016)
 - o Developed a analytics strategy for contextual recommendations and machine learning capabilities
 - o Directing efforts on proof of concept implementation
- [Vistaprint](#) (June 2016 - Dec 2016)
 - o Developed a context aware personalized recommendation engine for email campaigns
 - Included feature engineering and necessary code modification to function on a distributed (Spark) system.
 - Proof of concept model deployed to production, showed a 5-7% statistically significant lift in conversions
 - o General guidance on machine learning testing/evaluation.
- [Precision Nutrition](#) (July 2016 - Dec 2016, Sept 2017 - Present)
 - o Research and development focused on intervention strategies to reduce client churn.
 - o Initial results showed high accuracies (75+%) for predicting churn.
 - o Currently expanding research efforts to deliver deeper insights into habits/behaviors leading to churn (cancellation) and providing coaches with recommendations for intervention strategies.
 - These specifically focus on modeling and understanding factors leading to client success (e.g. weight loss) or failure.
 - Implemented an NLP model analyzing coach-client correspondences and cancellations using Doc2Vec (achieved predicted churn accuracy of 93%)
 - o Deployed these tools to Tableau, currently being used by PN coaches
- [ADEXT](#) (Feb 2017 - Present)
 - o Budget attribution modeling to determine spend on Facebook and Google platforms.
 - Developed visualization of campaign performance
 - Developed and implemented a multi-task gaussian process regression based approaches combined with Thompson Sampling to identify appropriate allocations for advertising spend
 - Production model has shown consistent 100% lift over baseline since production release (May/June 2018)

- Combined with LIME to assess budget policy impact
 - o Leading and providing general guidance on data science, machine learning and data collection, structuring and architecture.
 - o Implementation of trend detection models (identify those users who's conversion rates are showing decreasing trends, provide automated warning to CSR).
 - o Exploration of real time bidding methods.
 - o Assisted with technical due diligence preparation.
 - Resulted in a successful Series A funding round
- [WiserTogether/World Bank](#) (March 2017 - Sep 2017)
 - o Developed methods/models for entity disambiguation between shippers and consignees in bill of lading (BOL) data. Approaches included active learning using multinomial logistic models for consignee matching and fuzzy string matching for country/address identification for shippers.
 - o Creation of backend for BOL data aggregation/summarization.
- [Biostrap](#) (Feb 2018 - Present)
 - o Developing insights/recommendation engine for identifying factors impacting various biometrics (e.g. sleep quality, heart rate variability, heart rate, respiratory rate, and others).
 - Approach is based on a combination of ensemble models, gaussian processes and leveraging the LIME (locally interpretable model explanation) and SHAP frameworks.
 - Based on findings the goal is to recommend courses of action a user can take to optimize biometric measurements.
 - o Working with the Biostrap team on improving classification accuracy of complex movement patters (e.g. rather than simply walking or standing/sitting, exercises like deadlifts, clean and jerks, etc).
 - Approaches to be explored will leverage combining data across users normalized by gender, body proportions (i.e. arm, leg, torso length, etc) and various neural network based approaches (e.g. LSTM+CNN).
- [Aurora North Software](#) (August 2018 - Present)
 - o Developed text classification model for 400 million patent application to predict art unit classification
 - o Model developed achieved a 92% accuracy using top 3 recommended art units
 - o Beginning work on a project classifying email correspondences into predefined categories based on email text, PDF documents as well as a variety of metadata.
 - Development and implementation of name entity recognition models using Word2Vec and Doc2Vec.

- **Policy Engine**
 - o Providing general guidance and advice on data science and machine learning strategy.
- **Moen**
 - o Analyzed user shower usage and identified customer segments
 - o Recommended follow up data to be collected and potential insights customers might find valuable
- **Keelo**
 - o Implemented a proof of concept model for pose estimation to assist in analyzing weightlifter technique
- **Boston Consulting Group Digital Ventures**
 - o Providing general guidance in the development of a microsimulation platform for the biopharma space built on top of publicly available data (IHME, NHANES, etc.) as well as patient level data from commercial vendors.
 - o Intermediary between Boston Consulting Group, microsimulation subject matter experts and product managers.
 - o Exploration of automated calibration for microsimulations using Bayesian Optimization

Programming languages and frameworks

- Strong background in R/S, Python, MATLAB
- Experience with Ruby, Rails, C/C++, Fortran, MySQL, postgresql, MongoDB, Redshift, AWS, Git
- Exposure to Hadoop, Spark, Javascript, HTML, CSS, C#

Software
Available
:

- **calib**: Provides basic S4 data structures and routines for calibration of bioassays (available at <http://www.r-project.org/>). With Haaland PD and McVey E.

In Development (available upon request):

- **fRODEO**: Fast local polynomial regression based on the RODEO algorithm (R package).
- **SPLASSO**: Spatial LASSO (R package).
- **sga**: Sparse group analysis. Provides functionality for sparse regression with ridge and spatial $l_{\infty 2}$

regularization (R/C++).

- trecap: Analysis of TREC system performance using a bootstrap ANOVA (R and Python code available). *Git Resources*:

- <https://gitlab.com/u/dvsstats>
- <https://github.com/DanSBS>

Panels, Committees and Working Groups

- NIST Big Data Working Group (2011-2016). Website: <http://bigdatawg.nist.gov/home.php>.
- NIST/FDA workshop on Biosimilars, FDA, Silver Spring, MD (2012): “Panel: Applications of Mass Spectrometry for Characterizing Biopharmaceuticals in the Regulatory Setting”.
- Scientific Data Committee (2013-2014): Tasked with developing and implementing a roadmap for NIST to be in compliance with the open data initiative mandate. Serving on the data access and outreach subcommittees.

Conference

organization/committees/chairs

- 16th Annual International Conference on Intelligent Systems for Molecular Biology, Toronto, Canada (2008): Member of organizing committee and fundraising chair for International Society for Computational Biology Student Council Symposium.
- Joint Research Conference, NIST, Gaithersburg, MD (2010): Member of organizing committee and session chair.
- UseR, NIST, Gaithersburg, MD (2010): Member of organizing committee, session chair, responsible for invited speaker slides.
- Joint Statistical Meeting, Vancouver, BC (2010): Metrology session chair.
- Big Data Workshop, NIST, Gaithersburg, MD (2012): Member of organizing committee.

Reviewe

r

- Editorial Review Board (ERB): reviewer for 19 NIST research publications.
- NIST grant applications: reviewer for 9 application proposals.
- Methodology and Computing in Applied Probability
- Journal of Statistical Software

- Journal of Statistical Planning and Inference
- IEEE Transactions on Geoscience and Remote Sensing
- IEEE International Conference on Big Data
- Journal of Statistics Education
- Journal of the American Statistical Association
- Journal of Computational and Graphical Statistics
- IEEE Big Data Conference
- IEEE Journal of Biomedical and Health Informatics

Honors and Awards

- Awarded the Judson C French Award (2017) for work in the development of standards for next generation sequencing.
- Innovations in Measurement Science (co-PI): Living Genetic Sensor Foundry, \$6m funding over 5 years.
- National Institute of Standards and Technology, Building the Future Grant: Development of methods for assessing the performance of information retrieval algorithms, \$60,000 (co-PI) 2015-2016.
- Innovations in Measurement Science 2017 Grant (\$6m) “Living Genetic Sensor Foundry for Predictive Engineering of Biology”, Ross D, Paralanov V, Strychalski E, Kelman Z, Kusne G, Samarov DV.
- NIH Trainee in Bioinformatics and Computational Biology, UNC, Chapel Hill, NC (2006-2009).
- Becton Dickinson Travel Award, Becton Dickinson, Research Triangle Park, NC (2007).
- Scholars for Tomorrow Interdisciplinary Fellowship, UNC, Chapel Hill, NC (2004): Awarded to first year graduate students pursuing interdisciplinary studies.
- Burian Award for Academic Excellence, Rutgers College, NJ (2002): Awarded to graduating member of the varsity swim team with the highest cumulative GPA.
- Meyer Award for Leadership, Rutgers College, NJ (2002): Awarded to member of the varsity swim team showing outstanding leadership.
- Captain Men’s Varsity Swim Team, NCAA Division I, Rutgers College, NJ (2001 – 2002).
- Varsity Swim Team, NCAA Division I, Rutgers College, NJ (1998-2002).

Activities and Interests

- Top 3 finisher in various CrossFit competitions (2010-2013).
- CrossFit Games Mid-Atlantic Regionals Competitor (2010-2011, 2015).
- Completed US Marine Corps Marathon, Washington, DC (2008).

- Part of relay swimming across the English Channel (2007).
- Graduate School Student Senate, UNC, Chapel Hill, NC (2007-2008).
- Division I swimming at Rutgers University (1998-2002).
- Men's swimming team captain (2001-2002).
- Rock climbing.