

Alexej Gossmann

Web: alexejgossmann.com | LinkedIn: alexejgossmann | Github: agisga | Email: agossman@tulane.edu

AREAS OF INTEREST

Development of statistical and machine learning techniques for feature selection and prediction on big high-dimensional datasets, with focus on understanding how false research findings can arise, and how they can be avoided; applications in genomics and neuroimaging, and other medical applications of machine learning; translation of the statistical and machine learning methodology into a useful (software) product.

SKILLS

Core: Data science, statistics, machine learning, (applied) mathematics, scientific research.
Software, programming, tools: R, Python, Ruby, C, C++, Matlab, L^AT_EX, Linux/Unix, Git and Github, HTML, CSS, Vim, AWS and Google Cloud.
Domain knowledge: Genomics and medical imaging.
Language knowledge: German, Russian, English.

EDUCATION

PhD, Bioinnovation
Tulane University, New Orleans, Louisiana, August 2018
GPA: 3.967

Doctoral research, Mathematics (PhD candidacy, all but dissertation, withdrawn)
Tulane University, New Orleans, Louisiana, through January 2017
GPA: 3.978

MS, Statistics
Tulane University, New Orleans, Louisiana, May 2014
GPA: 3.975

BS, Mathematics
Technische Universität Darmstadt, Darmstadt, Germany, May 2012
GPA: 3.7 (1.54 German grade)

Workshops and summer schools
IPAM New Deep Learning Techniques, Los Angeles, CA, February 2018. | SAMSI Distributed and Parallel Data Analysis Workshop, September 2016. | SAMSI Optimization Opening Workshop, August 2016. | SAMSI Optimization Summer School, August 2016. | 2nd Summer Institute in Statistics for Big Data, University of Washington, Seattle, June 2016. | 21st Summer Institute in Statistical Genetics, University of Washington, Seattle, WA, June 2016. | 20th Summer Institute in Statistical Genetics, University of Washington, Seattle, WA, June 2015. | SAMSI Industrial Mathematical and Statistical Modeling Workshop for Graduate Students, June 2014.

EXPERIENCE

Staff Fellow, Mathematical Statistician	Division of Imaging, Diagnostics, and Software Reliability (CDRH/OSEL/DIDSR) at the U.S. Food & Drug Administration
FDA, Silver Spring, MD	Sep 2018 – Present
Scientific research and regulatory review responsibilities supporting the evaluation of products with a “Big Data” element, including machine learning and artificial intelligence (AI) algorithms when used on large biomedical datasets.	

Research assistantship (doctoral research)

The Multiscale Bioimaging and
Bioinformatics Laboratory

Tulane University, New Orleans, LA

Jan 2015 – Jul 2018

Research in statistics and machine learning with application in genomics and neuroimaging under the supervision of Dr. Yu-Ping Wang, resulting in six peer-reviewed publications (Gossmann et al. [2015], Cao et al. [2015, 2016], Gossmann et al. [2018a,c], Brzyski et al. [2018]), presentations at multiple conferences and workshops, and several open source software packages (developed in R and C++).

Student intern (ORISE)

Division of Imaging, Diagnostics, and
Software Reliability (CDRH/OSEL/DIDSR)
at the U.S. Food & Drug Administration

FDA, Silver Spring, MD

May – Aug 2017, and Jan – Feb 2018

Machine learning research (with software implementation in R) resulting in a conference presentation and associated publication Gossmann et al. [2018b] related to the evaluation of machine learning systems in medicine.

Student developer

Google Summer of Code 2015

Remote position

May – August 2015

Project: Adding Linear Mixed Effects Models Support to SciRuby (supervised by Piotr Prins from the Ruby Science Foundation). The created statistical software package (implemented in Ruby) received about 2700 downloads (https://rubygems.org/gems/mixed_models), and was considered for deployment at <http://www.genenetwork.org>.

Research assistantship

Department of Mathematics

Tulane University, New Orleans, LA

Jun – Aug 2014

Worked under the supervision of Dr. Oleksandr Gromenko on statistical methods for the analysis of spatio-temporal processes with software implementation in R and C++, and with application in weather prediction (results unpublished).

Statistics MS research project

Department of Mathematics

Tulane University, New Orleans, LA

Sept 2013 – May 2014

Analysis of bone growth data using mixed-effects smoothing spline ANOVA methods (supervised by Dr. Michelle Lacey) with data analyses performed in R, resulting in a journal publication Sammarco et al. [2015].

Academic Mentor

Tulane University, New Orleans, LA

Tulane University, New Orleans, LA

Oct 2017 – May 2018

Design, mentoring, and guidance of an undergraduate research project applying machine learning methods to a large neuroimaging-genomics dataset, resulting in two presentations by the mentored undergraduate student, entitled “*Exploratory Analysis and Predictive Modeling of Neurodevelopmental Phenotypes from fMRI Data*” at the 2018 Health Sciences Research Days at Tulane University, and at the 2018 Tulane School of Science and Engineering Poster Days (top 3 finalist in the poster competition).

Co-Mentor

Google Summer of Code 2016 (with the
Ruby Science Foundation)

Remote position

May – August 2016

Involved with project proposals, selection/interviewing of students, mentoring. Mentored project: Categorical data support for Daru, Statsample and Statsample-glm. The mentored student (Lokesh Sharma) made major improvements to several open source software packages for data analysis in Ruby.

Co-Mentor

Google Summer of Code 2017 (with the
Ruby Science Foundation)
May – August 2017

Remote position

Involved with project proposals, selection/interviewing of students, mentoring. Mentored project: Creating the fastest math libraries for Ruby by using the GPU through OpenCL and ArrayFire. Outcome: a Ruby library for scientific computing on the GPU (<https://github.com/arrayfire/arrayfire-rb>) and multiple conference presentations by the mentored student (Prasun Anand).

Instructor, co-teacher, teaching assistant

Technische Universität Darmstadt, and
Tulane University

Darmstadt, Germany, and New Orleans, LA

Fall semester 2010 – Fall semester 2014

Instructor for Calculus 1, co-teacher for Statistics for Scientists, and teaching assistant for Real Analysis 1 and 2 as well as for various undergraduate courses in mathematics and statistics.

PUBLICATIONS (sorted alphabetically)

[Brzyski et al., 2018] Damian Brzyski, Alexej Gossmann, Weijie Su, and Małgorzata Bogdan. Group SLOPE – Adaptive Selection of Groups of Predictors. *Journal of the American Statistical Association*, pages 1–15, January 2018. ISSN 0162-1459. doi:10.1080/01621459.2017.1411269.

[Cao et al., 2015] Shaolong Cao, Huaizhen Qin, Alexej Gossmann, Hong-Wen Deng, and Yu-Ping Wang. Unified tests for fine scale mapping and identifying sparse high-dimensional sequence associations. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '15, pages 241–249, New York, NY, USA, 2015. ACM. doi:10.1145/2808719.2808744.

[Cao et al., 2016] Shaolong Cao, Huaizhen Qin, Alexej Gossmann, Hong-Wen Deng, and Yu-Ping Wang. Unified tests for fine-scale mapping and identifying sparse high-dimensional sequence associations. *Bioinformatics*, 32(3):330–337, February 2016. doi:10.1093/bioinformatics/btv586.

[Gossmann et al., 2015] Alexej Gossmann, Shaolong Cao, and Yu-Ping Wang. Identification of significant genetic variants via SLOPE, and its extension to Group SLOPE. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '15, pages 232–240, New York, NY, USA, 2015. ACM. doi:10.1145/2808719.2808743.

[Gossmann et al., 2018a] Alexej Gossmann, Shaolong Cao, Damian Brzyski, Lan-Juan Zhao, Hong-Wen Deng, and Yu-Ping Wang. A sparse regression method for group-wise feature selection with false discovery rate control. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(4):1066–1078, July 2018a. ISSN 1545-5963, 1557-9964. doi:10.1109/TCBB.2017.2780106.

[Gossmann et al., 2018b] Alexej Gossmann, Aria Pezeshk, and Berkman Sahiner. Test data reuse for evaluation of adaptive machine learning algorithms: Over-fitting to a fixed “test” dataset and a potential solution. In *Proceedings of SPIE: Medical Imaging 2018*. International Society for Optics and Photonics, March 2018b. doi:10.1117/12.2293818.

[Gossmann et al., 2018c] Alexej Gossmann, Pascal Zille, Vince Calhoun, and Yu-Ping Wang. FDR-Corrected Sparse Canonical Correlation Analysis with Applications to Imaging Genomics. *IEEE Transactions on Medical Imaging*, 37(8):1761–1774, August 2018c. ISSN 0278-0062, 1558-254X. doi:10.1109/TMI.2018.2815583.

[Gossmann et al., 2019] Alexej Gossmann, Kenny H. Cha, and Xudong Sun. Variational inference based assessment of mammographic lesion classification algorithms under distribution shift. In *Medical Imaging Meets NeurIPS Workshop (MED-NeurIPS) 2019*, De-

cember 2019. URL https://profs.etsmtl.ca/hlombaert/public/medneurips2019/72_CameraReadySubmission_neurips_2019.pdf.

[Hosseinzadeh Kassani et al., 2019] Peyman Hosseinzadeh Kassani, Alexej Gossmann, and Yu-Ping Wang. Multimodal Sparse Classifier for Adolescent Brain Age Prediction. *IEEE journal of biomedical and health informatics*, June 2019. ISSN 2168-2208, 2168-2194. doi:10.1109/JBHI.2019.2925710. URL <http://dx.doi.org/10.1109/JBHI.2019.2925710>.

[Sammarco et al., 2015] Mimi C Sammarco, Jennifer Simkin, Alexander J Cammack, Danielle Fassler, Alexej Gossmann, Luis Marrero, Michelle Lacey, Keith Van Meter, and Ken Muneoka. Hyperbaric oxygen promotes proximal bone regeneration and organized collagen composition during digit regeneration. *PloS one*, 10(10), 2015. doi:10.1371/journal.pone.0140156.

[Sun et al., 2019] X Sun, A Gossmann, Y Wang, and B Bischt. Variational Resampling Based Assessment of Deep Neural Networks under Distribution Shift. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1344–1353, December 2019. doi:10.1109/SSCI44817.2019.9002665. URL <http://dx.doi.org/10.1109/SSCI44817.2019.9002665>.

CONFERENCE PRESENTATIONS

- Gossmann A. Evaluation of continuously learning AI/ML-enabled medical devices. Presented at *Next Generation Dx*, August 2019, Washington, DC.
- Gossmann A., Chen W., Sahiner, B., Assessment of Classifier Performance Using a Reference Classifier with Known Performance and an Unlabeled Dataset; (Abstract/Program #307008). Presented at the *Joint Statistical Meetings*, July 28, 2019, Denver, CO.
- Gossmann A., Pezeshk, A., Sahiner, B., Test data reuse for evaluation of adaptive machine learning algorithms: over-fitting to a fixed “test” dataset and a potential solution; (Paper 10577-19). Presented at the *SPIE Medical Imaging symposium, Image Perception, Observer Performance, and Technology Assessment conference*, February 11, 2018, Houston, TX.
- Gossmann A., Wang Y.-P., SlopeCCA and gslopeCCA: sorted L1 penalized canonical correlation analysis; (Abstract/Program #2803W). Presented at the *66th Annual Meeting of The American Society of Human Genetics*, October 19, 2016, Vancouver, Canada.
- Gossmann A., Cao S., Wang Y.-P., Identification of Significant Genetic Variants via SLOPE, and its Extension to Group SLOPE; (Abstract/Program #1343F). Presented at the *65th Annual Meeting of The American Society of Human Genetics*, October 9, 2015, Baltimore, MD.
- Gossmann A., Cao S., Wang Y.-P., Identification of Significant Genetic Variants via SLOPE, and its Extension to Group SLOPE. The *6th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, September 2015, Atlanta, GA.

SOFTWARE

- **grpSLOPE** – Group SLOPE (Group Sorted L1 Penalized Estimation) is a penalized linear regression method that is used for adaptive selection of groups of significant predictors in a high-dimensional linear model. This R package has been used to perform simulations and/or analyses of real genomic data as reported in Gossmann et al. [2018a], Brzyski et al. [2018]. Project repository: <https://github.com/agisga/grpSLOPE>. CRAN: <https://cran.r-project.org/web/packages/grpSLOPE/index.html>.
- **grpSLOPEMC** – An extension package to the R package grpSLOPE, which contains additional Monte Carlo based methods implemented in R and C++ (interfaced to R

via Rcpp). This R package has been used to perform simulation studies and analyses of real genomic data presented in Gossmann et al. [2018a]. Project repository: <https://github.com/agisga/grpSLOPEMC>.

- **FDRcorrectedSCCA** – Codes associated with the publication Gossmann et al. [2018c] with all methods and algorithms conveniently organized as functions in an R package. Project repository: <https://github.com/agisga/FDRcorrectedSCCA>.
- **mixed_models** – Fit statistical linear models with fixed and random effects in Ruby. It was created during Google Summer of Code 2015, and has been considered for deployment at <http://www.genenetwork.org>. Project repository: https://github.com/agisga/mixed_models. RubyGems: https://rubygems.org/gems/mixed_models.
- For my contributions to several other open source software projects see <https://github.com/agisga>.

EXTRA-CURRICULAR ACTIVITIES

- Elected *President* of the SIAM student chapter at Tulane University (Society for Industrial and Applied Mathematics), September 2014 – September 2016.
- Organized the Graduate Student Colloquium at the Department of Mathematics, Tulane University, September 2014 – September 2016.
- Participation in *Google Summer of Code* as a student developer in 2015, and as a mentor in 2016 and 2017, working on scientific open source software.
- Participation in the middle school outreach program *Girls in STEM at Tulane (GIST)*, November 2017.

HONORS & AWARDS

- Finalist poster at the Tulane School of Science and Engineering Poster Days 2018.
- Reviewers' Choice Abstract, 66th Annual Meeting of The American Society of Human Genetics, October 19, 2016, Vancouver, Canada.
- SIAM Student Chapter Certificate of Recognition, SIAM (Society for Industrial and Applied Mathematics), April 2015.