

## 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 positive 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.

### EDUCATION

*PhD*, Bioinnovation

Tulane University, New Orleans, Louisiana, in progress

*Doctoral research*, Mathematics

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)

### SKILLS

*Software, programming, tools*: R, Ruby, C++, Python, Matlab, L<sup>A</sup>T<sub>E</sub>X, Linux/Unix, git and github, HTML, CSS.

*Domain knowledge*: Genomics and medical imaging.

*Language knowledge*: Bilingual in German and Russian, fluent in English.

### EXPERIENCE

Research assistantship (doctoral research)

The Multiscale Bioimaging and  
Bioinformatics Laboratory

Tulane University, New Orleans, LA

Jan 2015 – present

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., 2015a], [Cao et al., 2015b], [Gossmann et al., 2017a], [Gossmann et al., 2017b], [Brzyski et al., 2018]), presentations at multiple conferences and workshops, and several open source software packages.

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 resulting in a conference presentation and associated publication [Gossmann et al., 2018] that can inform future policy on evaluation of machine learning systems in medicine.

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) resulting in a journal publication [Sammarco et al., 2015].

Research assistantship  
Tulane University, New Orleans, LA  
Department of Mathematics  
Jun – Aug 2014  
Worked under the supervision of Dr. Oleksandr Gromenko on statistical methods for the analysis of spatio-temporal processes with application in weather prediction.

Student developer  
Remote position  
Google Summer of Code 2015  
May – August 2015  
Project: Adding Linear Mixed Effects Models Support to SciRuby (supervised by developers from the Ruby Science Foundation). The created statistical software package received about 2700 downloads, and was considered for deployment at <http://www.genenetwork.org>.

Mentor  
Remote position  
Google Summer of Code 2016 (with the Ruby Science Foundation)  
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.

Mentor  
Remote position  
Google Summer of Code 2017 (with the Ruby Science Foundation)  
May – August 2017  
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  
Darmstadt, Germany, and New Orleans, LA  
Technische Universität Darmstadt, and Tulane University  
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

- [Brzyski et al., 2018] Brzyski, D., Gossmann, A., Su, W., and Bogdan, M. (2018). Group SLOPE-adaptive selection of groups of predictors. *Accepted for publication in the Journal of the American Statistical Association Theory and Methods Section*.
- [Cao et al., 2015a] Cao, S., Qin, H., Gossmann, A., Deng, H.-W., and Wang, Y.-P. (2015a). 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. ACM.
- [Cao et al., 2015b] Cao, S., Qin, H., Gossmann, A., Deng, H.-W., and Wang, Y.-P. (2015b). Unified tests for fine scale mapping and identifying sparse high-dimensional sequence associations. *Bioinformatics*.
- [Gossmann et al., 2017a] Gossmann, A., Cao, S., Brzyski, D., Zhao, L. J., Deng, H. W., and Wang, Y. P. (2017a). A sparse regression method for group-wise feature selection with false discovery rate control. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, PP(99):1–13.

- [Gossmann et al., 2015] Gossmann, A., Cao, S., and Wang, Y.-P. (2015). 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. ACM.
- [Gossmann et al., 2018] Gossmann, A., Pezeshk, A., and Sahiner, B. (2018). 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*.
- [Gossmann et al., 2017b] Gossmann, A., Zille, P., Calhoun, V., and Wang, Y.-P. (2017b). FDR-corrected sparse canonical correlation analysis with applications to imaging genomics. *arXiv preprint arXiv:1705.04312 (under review in IEEE/TMI)*.
- [Sammarco et al., 2015] Sammarco, M. C., Simkin, J., Cammack, A. J., Fassler, D., Gossmann, A., Marrero, L., Lacey, M., Van Meter, K., and Muneoka, K. (2015). Hyperbaric oxygen promotes proximal bone regeneration and organized collagen composition during digit regeneration. *PloS one*, 10(10).

## CONFERENCE PRESENTATIONS

- 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.
- 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.
- Identification of Significant Genetic Variants via SLOPE, and its Extension to Group SLOPE. The *6th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, Atlanta, GA, September 2015.

## 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. Group SLOPE can control the (group) false discovery rate at a user-specified level. Project repository: <https://github.com/agisga/grpSLOPE>. CRAN: <https://cran.r-project.org/web/packages/grpSLOPE/index.html>.
- **grpSLOPEMC** – This is an extension package to the R package grpSLOPE. It contains Monte Carlo based methods for the estimation of the regularizing sequence. Project repository: <https://github.com/agisga/grpSLOPEMC>.
- **mixed\_models** – Fit statistical linear models with fixed and random effects in Ruby. Project repository: [https://github.com/agisga/mixed\\_models](https://github.com/agisga/mixed_models). RubyGems: [https://rubygems.org/gems/mixed\\_models](https://rubygems.org/gems/mixed_models).
- In my free time I like to contribute to open source software projects. Visit my Github page for the projects that I contribute to: <https://github.com/agisga>.

## EXTRA-CURRICULAR ACTIVITIES

Elected *President* of the SIAM student chapter at Tulane University, 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. Participation in the middle school outreach program *Girls in STEM at Tulane (GIST)*, November 2017.

## **HONORS & AWARDS**

- Reviewers Choice Abstract (top 10% of poster abstracts), 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.