

Sangwon (Justin) Hyun

Email: sangwonh@ucsc.edu

Github: <https://github.com/sangwon-hyun/>

Website: <https://sangwon-hyun.org>

RESEARCH INTERESTS

Keywords Oceanographic data analysis
Changepoint detection
Selective inference
Applied optimal transport
Flow cytometry data analysis
Infectious disease modeling and forecasting

EDUCATION

2018 Ph.D. in Statistics, *Carnegie Mellon University, Pittsburgh PA*
Thesis: Post-selection inference for changepoint-type problems
Advisors: Ryan Tibshirani, Max G'Sell

2014 M.S. in Statistics, *Carnegie Mellon University, Pittsburgh PA*

2013 B.S. in Mathematics and Honors Statistics, *University of Michigan, Ann Arbor MI*
Advisor: Ed Rothman

POSITIONS HELD

07/2022 – present Assistant Professor
Department of Statistics, University of California, Santa Cruz

01/2019 – 06/2022 Post-doctoral Researcher
Department of Data Sciences and Operations, University of Southern California
Advisor: Jacob Bien

PUBLICATIONS (PUBLISHED OR UNDER REVIEW)

Selective Inference and Changepoint Detection

Sangwon Hyun, Kevin Z. Lin, Max G'Sell, and Ryan J. Tibshirani. Post-selection inference for changepoint detection algorithms with application to copy number variation data. *Biometrics*, 2021.

Sangwon Hyun, Max G'Sell, and Ryan J. Tibshirani. Exact post-selection inference for the generalized lasso path. *Electronic Journal of Statistics*, 2018.

Ocean Data Analysis

- Sangwon Hyun, Aditya Mishra, Christopher L. Follett Bror Jonsson, Gemma Kulk, Gael Forget, Marie-Fanny Racault, Thomas Jackson, Stephanie Dutkiewicz, Christian L. Müller, and Jacob Bien. Ocean mover's distance: Using optimal transport for analyzing oceanographic data. *Proceedings of the Royal Society A: Mathematical, Physical*, 2022. URL <https://royalsocietypublishing.org/doi/full/10.1098/rspa.2021.0875>.
- Jann Paul Mattern, Kristof Glauning, Gregory L. Britten, John Casey, Sangwon Hyun, Zhen Wu, E Virginia Armbrust, Zaid Harchaoui, and François Ribalet. A flexible Bayesian approach to estimating size-structured matrix population models. *PLOS Computational Biology*, 2022. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009733>.
- Bror F. Jönsson, Christopher L. Follett, Jacob Bien, Stephanie Dutkiewicz, Sangwon Hyun, Gemma Kulk, Gael Forget, Christian L. Müller, Marie-Fanny Racault, Christopher N. Hill, Thomas Jackson, and Shubha Sathyendranath. Density-based approaches to evaluate biogeochemical models by satellite derived chlorophyll. *In revision, AGU Global Biogeochemical Cycles*, 2021.
- Sangwon Hyun, Mattias Rolf Cape, François Ribalet, and Jacob Bien. Modeling cell populations measured by flow cytometry with covariates using sparse mixture of regressions. *To appear, Annals of Applied Statistics*, 2020. URL <https://arxiv.org/abs/2008.11251>.

Infectious Disease Forecasting

- Daniel J. McDonald, Jacob Bien, Alden Green, Addison J. Hu, Nat DeFries, Sangwon Hyun, Natalia L. Oliveira, James Sharpnack, Jingjing Tang, Robert Tibshirani, Valérie Ventura, Larry Wasserman, and Ryan J. Tibshirani. Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction? *To appear, Proceedings of the National Academy of Sciences*, 2021. URL <https://www.medrxiv.org/content/10.1101/2021.06.22.21259346v1>.
- Alex Reinhart, Logan Brooks, Maria Jahja, Aaron Rumack, Jingjing Tang, Wael Al Saeed, Taylor Arnold, Amartya Basu, Jacob Bien, Ángel A. Cabrera, Andrew Chin, Eu Jing Chua, Brian Clark, Nat DeFries, Jodi Forlizzi, Samuel Gratzl, Alden Green, George Haff, Robin Han, Addison J. Hu, Sangwon Hyun, Ananya Joshi, Jimi Kim, Andrew Kuznetsov, Wichada La Motte-Kerr, Yeon Jin Lee, Kenneth Lee, Zachary C. Lipton, Michael X. Liu, Lester Mackey, Kathryn Mazaitis, Daniel J. McDonald, Balasubramanian Narasimhan, Natalia L. Oliveira, Pratik Patil, Adam Perer, Collin A. Politsch, Samyak Rajanala, Dawn Rucker, Nigam H. Shah, Vishnu Shankar, James Sharpnack, Dmitry Shemetov, Noah Simon, Vishakha Srivastava, Shuyi Tan, Robert Tibshirani, Elena Tuzhilina, Ana Karina Van Nortwick, Valérie Ventura, Larry Wasserman, Jeremy C. Weiss, Kristin Williams, Roni Rosenfeld, and Ryan J. Tibshirani. An open repository of real-time COVID-19 indicators. *To appear, Proceedings of the National Academy of Sciences*, 2021. URL <https://www.medrxiv.org/content/10.1101/2021.06.22.21259346v1>.
- Craig J McGowan, Matthew Biggerstaff, Michael Johansson, Karyn M Apfeldorf, Michal Ben-Nun, Logan Brooks, Matteo Convertino, Madhav Erraguntla, David C Farrow, John Freeze, et al. Collaborative efforts to forecast seasonal influenza in the United States, 2015–2016. *Scientific reports*, 2019.
- Michael A Johansson, Karyn M Apfeldorf, Scott Dobson, Jason Devita, Anna L Buczak, Benjamin Baugher, Linda J Moniz, Thomas Bagley, Steven M Babin, Erhan Guven, et al. An open challenge to advance probabilistic forecasting for Dengue epidemics. *Proceedings of the National Academy of Sciences*, 2019.
- Matthew Biggerstaff, Michael Johansson, David Alper, Logan C Brooks, Prithwish Chakraborty, David C Farrow, Sangwon Hyun, Sasikiran Kandula, Craig McGowan, Naren Ramakrishnan, et al. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*, 2018.
- Logan C. Brooks, David C. Farrow, Sangwon Hyun, Ryan J. Tibshirani, and Roni Rosenfeld. Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions. *PLOS Computational Biology*, 2018.

David C. Farrow, Logan C. Brooks, Sangwon Hyun, Ryan J. Tibshirani, Donald S. Burke, and Roni Rosenfeld. A human judgment approach to epidemiological forecasting. *PLOS Computational Biology*, 2017.

Logan C. Brooks, David C. Farrow, Sangwon Hyun, Ryan J. Tibshirani, and Roni Rosenfeld. Flexible modeling of epidemics with an Empirical Bayes framework. *PLOS Computational Biology*, 2015.

Willem G. van Panhuis, Sangwon Hyun, Kayleigh Blaney, Ernesto T. A. Marques, Jr, Giovanini E. Coelho, João Bosco Siqueira, Jr, Ryan Tibshirani, Jarbas B. da Silva, Jr, and Roni Rosenfeld. Risk of Dengue for tourists and teams during the World Cup 2014 in Brazil. *PLOS Neglected Tropical Diseases*, 2014.

WORK IN PROGRESS

Sangwon Hyun, Timothy Coleman, Francois Ribalet, and Jacob Bien. Environmental drivers of marine cell populations measured by flow cytometry. *In preparation*, 2021a.

Timothy Coleman, Sangwon Hyun, Francois Ribalet, and Jacob Bien. Trend filtering for mixture analysis of flow cytometry. *In preparation*, 2021.

Ryan Reynolds, Sangwon Hyun, Jacob Bien, and Naomi Levine. Using aspect Bernoulli matrix decomposition for analyzing metagenomic data from the ocean. *In preparation*, 2021.

Sangwon Hyun, Natalia Lombardi de Oliveira, and Lester Mackey. Estimating missing COVID-19 prevalence from massive digital surveys. *In preparation*, 2021b.

ACTIVE COLLABORATIONS

- [Armbrust Lab](#), School of Oceanography, University of Washington, WA.
- [Levine Lab](#), Marine and Environmental Biology, University of Southern California, CA.
- [Plymouth Marine Laboratory](#), Plymouth, UK.
- [Follows Lab](#) and the [Darwin Project](#) at the Massachusetts Institute of Technology, MA.
- [Müller group](#), Ludwig Maximilian University of Munich, Germany.
- [DELPHI group](#), Carnegie Mellon University, PA.

AWARDS

- 2021 ASA Statistical Partnerships Among Academe, Industry, and Government (SPAIG) Award
 - Group award to CMU DELPHI group.
 - Awarded for outstanding partnerships between academe, industry, and government organizations.
- (2014-2018) Winner, CDC Predict the Influenza challenge
 - Group award to CMU DELPHI group.
 - Most accurate forecasting of seasonal influenza, four years in a row.
- 2016 Best student poster, American Association for Advancement of Science (AAAS)
 - Title: [Epidemiological forecasting with statistical models](#)
- 2012 Winner, Yahoo! Data Mining Competition
 - Awarded by University of Michigan Informatics & Statistics department.

POSTERS AND PRESENTATIONS

Invited talk: Sparse Gaussian mixture regression with application to flow cytometry data analysis. *International Chinese Statistical Association (ICSA) Conference*, Virtual, 2022.

Invited talk: Nonconvex, nonsmooth, constrained optimization for analyzing the ocean's microbial communities. *INFORMS Optimization Society Conference*, Greenville, SC, 2022.

Ocean mover's distance: flexible comparison of probability distributions across the ocean optimal transport for analyzing ocean data. *The 1st International Workshop on Optimal Transport and Structured Data Modeling*, Co-organized with AAAI 2022, Virtual, 2022.

Invited plenary talk: Ocean mover's distance: flexible comparison of probability distributions across the ocean optimal transport for analyzing ocean data. *Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems (CBIOMES) Annual Meeting*, Virtual, 2021.

Optimal transport for analyzing ocean data. *Joint Statistical Meetings*, Seattle, WA, 2021.

Examining leading indicators for COVID-19 incidence case growth. *Symposium on Data Science and Statistics*, Virtual, 2021.

Invited talk: High-performance computing with R: a case study in ocean flow cytometry analysis. *(Delayed) USC Advanced Research Computing Colloquium*, Los Angeles, CA, 2021.

Invited talk: Sparse Gaussian mixture regression with application to flow cytometry data analysis. *4th International Conference on Econometrics and Statistics*, Virtual, 2021.

Identifying ocean cell population mixtures using environmental factors. Poster, *American Association for Advancement of Science (AAAS) annual meeting*, Virtual, 2020.

Invited talk: Sparse multivariate mixture of regressions modeling for flow cytometry data. *Flatiron Institute Center for Computational Mathematics Seminar*, New York, NY, 2020.

Joint modeling of continuous flow cytometry data with environmental covariates. *Joint Statistical Meetings*, Philadelphia, PA, 2020.

Invited talk: Joint modeling of continuous flow cytometry data with environmental covariates. *(Canceled) Statistical Learning and Data Science*, Newport Beach, CA, 2020.

Determining ecological provinces from optical cytometric data in the north pacific ocean. *(Presented by co-author) Ocean Sciences Meeting*, San Diego, CA, 2020.

Joint modeling of continuous flow cytometry data with environmental covariates. *CMStatistics*, London, UK, 2019.

Knockoff variable selection for changepoint detection. Poster, *Symposium on Data Science and Statistics*, Reston, VA, 2018.

On changepoint inference using binary segmentation inference. *Joint Statistical Meetings*, Baltimore, MD, 2017.

Invited talk: Forecasting of Dengue risk in 2016 for Southeast Asia. *Southeast Asia regional meeting on climate and dengue forecasting* Kuala Lumpur, Malaysia, 2016.

Epidemiological forecasting with statistical models. Best Student Poster, *American Association for Advancement of Science (AAAS) annual meeting*, Washington DC, 2016.

On changepoint inference after selection. *Joint Statistical Meetings*, Chicago, IL, 2016.

On changepoint inference after selection. *INFORMS*, Nashville, TN, 2016.

NON-ACADEMIC WORK EXPERIENCE

06/2015 – 08/2015 Research intern

Langley NASA Research Center, Hampton, VA

09/2012 – 07/2013 Desk consultant

Center for Statistical Consulting And Research, University of Michigan, Ann Arbor, MI

09/2012 – 04/2013 Tutor

Statistics tutor, Comprehensive Studies Program, University of Michigan, Ann Arbor, MI

SOFTWARE

- flowmix: R package for sparse multivariate mixture of regressions
 - Link: <https://github.com/sangwon-hyun/flowmix>
- cmap4r: R package for accessing Simons CMAP ocean database
 - Link: <https://github.com/simonscmap/cmap4r>
- genlassoinf: R package for generalized lasso post-selection inference
 - Link: <https://github.com/sangwon-hyun/genlassoinf>
- binseginf: R package for segmentation post-selection inference
 - Link: <https://github.com/sangwon-hyun/binseginf>
- epiforecast: R package for epidemiological forecasting
 - Link: <https://github.com/cmu-delphi/epiforecast-R>

TEACHING EXPERIENCE

Full Instructor

- 36-220 Engineering Statistics and Quality Control
- 36-217 Probability Theory and Random Processes

Teaching Assistant

- 36-217 Probability theory and random processes
- 36-225 & 36-226 Mathematical statistics sequence for undergraduates
- 36-350 Undergraduate statistical computing
- 36-402 Undergraduate advanced data analysis
- 36-617 Applied linear models
- 36-618 Topics in statistics
- 36-725 Convex Optimization
- 36-750 Statistical Computing

Teaching Research

- I'm interested in how to improve teaching in statistics and data science. I was involved in the CMU statistics department's effort to revamp and improve the undergraduate curriculum, starting with introductory courses. I co-organized two seminar courses, both named 36-764, in which we discussed literature on learning, created a repository of assessment test questions (focused on testing conceptual understanding), and interviewed students in order to probe their misunderstanding and improve test questions. See more details in the group's website <http://stat.cmu.edu/teachstat/>.

Assessment of student learning and misconception identification in intro statistics. Poster presentation, *Eberly Teaching and Learning Summit*, Pittsburgh, PA, 2017.

Identifying misconceptions of introductory data science using a think-aloud protocol. Poster presentation, *eCOTS*, Pittsburgh, PA, 2018.

Student Supervision

- During my post-doc, I supervised an undergraduate student who built an interactive data visualization application (using R Shiny and Plotly) for the `flowmix` flow cytometry model.

PROFESSIONAL ACTIVITIES & SERVICE

General Service

- 2019-2021 Vice President, Post-doctoral Association, USC
- 2013-2018 Student Advisory Council, Statistics Department, CMU
- 2015-2018 Co-organizer, Student Seminar Series, Statistics Department, CMU
- 2016 Volunteer Judge for Undergraduate Tartan Data Science Cup, CMU
- 2016-2018 Member of Statistics Department Help Network, CMU

Community Service

- 2016-2018 Volunteer Korean teacher at Carnegie Public Library, Pittsburgh, PA
- 2016 Volunteer judge at K-12 Covestro Pittsburgh Regional Science & Engineering Fair, Pittsburgh, PA
- 2017 Volunteer judge at K-12 Covestro Pittsburgh Regional Science & Engineering Fair, Pittsburgh, PA
- 2019 Volunteer judge at K-12 Los Angeles County Science & Engineering Fair, Los Angeles, LA

Academic Referee Activity

- Journal of Machine Learning
- 20th and 21st International Conference on Artificial Intelligence and Statistics
- Bernoulli Journal
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Neural Processing Letters (NEPL)
- Journal of Statistical Planning and Inference
- Journal of Econometrics
- Biometrika
- Journal of the American Statistical Association
- Journal of the Royal Statistics Society Part B

- Statistical Sinica
- PLOS Neglected Tropical Diseases.
- BMC Infectious Diseases
- Electric Journal of Statistics

ADDITIONAL SKILLS

Computer **High performance computing** Proficient in using SLURM Workload Manager (served as beta user during upgrade [USC Advanced Research Computing](#))

Operating System Expert Linux user.

Programming Languages R (C++ through R); Python; \LaTeX .

Programs and OS Daily Emacs and Org-mode user, on an Ubuntu laptop.

Language **Korean** Native language

English Native language

OUTSIDE OF WORK

- I enjoy hiking and visit national parks whenever I get the chance. I like recreational weight-lifting and biking, and enjoy playing all sports that involve bouncy spheres (especially soccer).
- Fun fact: I served 2 years in the military at the [Joint Security Area](#), at the border of North/South Korea.