

Emily C. Hector

North Carolina State University

Email: ehector@ncsu.edu

Website: emilyhector.com

Department of Statistics

2311 Stinson Drive

5140 SAS Hall

Raleigh, NC

RESEARCH INTERESTS

Theory/Methods

Composite likelihood, Correlated data, Divide-and-conquer, Distributed estimation and inference, Estimating equations, Generalized method of moments, Heterogeneous data integration, High-dimensional data, Parallel computing.

Applications

Brain imaging analysis, Metabolomics, Spatial data, Wearable devices.

PROFESSIONAL POSITIONS

Assistant Professor, Department of Statistics 2020-present
North Carolina State University

Graduate Student Research Assistant, Department of Biostatistics 2015-2020
University of Michigan

EDUCATION

PhD Biostatistics 2020
University of Michigan

Thesis: "Distributed estimation and inference for the analysis of big biomedical data"

Advisor: Peter X.-K. Song, PhD

MSc Biostatistics 2016
University of Michigan

BSc Honors Probability and Statistics 2014
McGill University

AWARDS & HONORS

1. Faculty Research and Professional Development award, North Carolina State University (2021).
2. *Finalist*, grant proposal submitted to the *Second Joint Biostatistics-Statistics Research Retreat, Shark Tank for Research Ideas in Data Science and Statistics (STRIDES)*. Departments of Biostatistics and Statistics, University of Michigan (2020)
3. *Excellence in Research Award*, awarded annually to one student in recognition of research excellence. Department of Biostatistics, University of Michigan (2019)
4. *Gertrude M. Cox Scholarship, Honorable Mention*, sponsored by the American Statistical Association (ASA) Committee on Women in Statistics and the Caucus for Women in Statistics (CWS) (2019)
5. *John Van Ryzin Award* for most outstanding paper submitted to the International Biometric Society (IBS) Eastern North American Region's (ENAR) Distinguished Student Paper Award Competition (2018)
6. *Rackham Conference Travel Grant*, University of Michigan (2016, 2017, 2018, 2019)
7. *Outstanding First-Year Masters Student*, Department of Biostatistics, University of Michigan (2015)
8. *First Class Honors*, McGill University (2014)
9. *Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award (USRA)*, McGill University (2013)
10. *J W McConnell Scholarship* (major award), McGill University (2011-2013)

FUNDING

Current

Projecting flood frequency curves under a changing climate using spatial extreme value analysis (Reich)
NSF, Role: co-PI (2022-2025)

Development and Application of New Ionization Methods for Biological Mass Spectrometry (Muddiman)
NIH, Role: co-I (2022-2026)

Past

Functional regression for intensive longitudinal data: a new lens through data partitioning (Hector)

NCSU Faculty Research and Professional Development award, Role: PI (2021 - 2022)

PUBLICATIONS

* co-first author; † corresponding author; + student author

Peer-reviewed journal articles – Statistical Methodology

1. Shi L+, Wank M+, Chen Y+, Wang Y+, Liu Y, **Hector EC**, Song P XK†. Sleep Classification with Artificial Synthetic Imaging Data from Empatica E4 Wristband by Convolutional Neural Networks (2022). To appear in *IEEE Journal of Biomedical and Health Informatics*.
2. **Hector EC***, Luo L*, Song P XK. Parallel-and-stream accelerator for computationally fast supervised learning. *Computational Statistics and Data Analysis* (2022). 177:107587.
3. **Hector EC**†, Song P XK. Joint integrative analysis of multiple data sources with correlated vector outcomes. *The Annals of Applied Statistics* (2022). 16(3):1700-1717.
4. **Hector EC**†, Song P XK. Doubly distributed supervised learning and inference with high-dimensional correlated outcomes. *Journal of Machine Learning Research* (2020). 21(173):1–35.
5. **Hector EC**†, Song P XK†. A distributed and integrated method of moments for high-dimensional correlated data analysis. *Journal of the American Statistical Association* (2021). 116(534):805-818.

Peer-reviewed journal articles – Statistical Applications

1. Yin X†, Chan LS, Bose D, Jackson AU, VandeHaar P, Locke AE, Fuchsberger C, Stringham HM, Yu K, Fernandes Silva L, Zhang D, **Hector EC**, Young E, Ganel L, Das I, Abel HJ, Erdos MR, Bonnycastle LL, Kuusisto J, Stitzziel NO, Hall I, Wagner GR, Kang J, Morrison J, Burant CF, Collins FS, Ripatti S, Palotie A, Freimer NB, Mohlke KL, Scott L, Wen X, Fauman E, Laakso M, Boehnke M. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nature Communications* (2022), doi: 10.1038/s41467-022-29143-5.

2. Goodrich JM*†, **Hector EC***, Tang L, Labarre JL, Dolinoy DC, Mercado-Garcia A, Cantoral A, Song P XK, Téllez Rojo MM, Peterson KE. Integrative analysis of gene-specific DNA methylation and untargeted metabolomics data from the ELEMENT cohort. *Epigenetic Insights* (2020). 13:1-10. doi: 10.1177/2516865720977888.
3. Jansen EC†, **Hector EC**, Goodrich JM, Cantoral A, Téllez Rojo MM, Basu N, Song P XK, Torres Olascoaga L, Peterson KE. Mercury exposure in relation to sleep duration, timing, and fragmentation among adolescents in Mexico City. *Environmental Research* (2020). 191: 110216. doi: j.envres.2020.110216.
4. Perng W†, **Hector EC**, Song P XK, Tellez Rojo MM, Raskind S, Kachman M, Cantoral A, Burant BF, Peterson KE. Metabolomic determinants of metabolic risk in Mexican adolescents. *Obesity (Silver Spring)* (2017). doi:10.1002/oby.21926.

Book chapters with peer-review

1. **Hector EC**†, Tang L, Zhou L, Song P XK (2022+). Data integration and fusion in the Bayesian and Frequentist frameworks. Chapter in “Handbook on Bayesian, Fiducial and Frequentist Inference”, tentatively accepted. In press.

Journal articles in review or revision

1. Manschot C⁺⁺, **Hector EC** (2022+). Functional regression with intensively measured longitudinal outcomes: a new lens through data partitioning.
2. **Hector EC**†, Martin R (2022+). Turning the information-sharing dial: efficient inference from different data sources.
3. **Hector EC**†, Reich BJ (2022+). Distributed inference for spatial extremes modeling in high dimensions.
4. **Hector EC**† (2020+). Fused mean structure learning in data integration with dependence.
5. Luo L, Wang J and **Hector EC**† (2021+). Statistical inference for streamed longitudinal data.

SOFTWARE

R packages

1. BRdac: Divide-and-conquer estimation and inference for max-stable spatial process modeling.
2. SLA: Performs streaming inference of intensively measured longitudinal data.
3. DIQIF: Performs joint integrative regression analysis of multiple data sources with correlated vector outcomes using quadratic inference functions.

4. DDIMM: Performs doubly distributed and integrated method of moments regression for high-dimensional correlated responses.
5. DIMM: Performs singly distributed and integrated method of moments regression for high-dimensional correlated responses.

TEACHING

North Carolina State University, Department of Statistics

ST422: Introduction to Mathematical Statistics II	(Fall '21, '22)
ST502: Fundamentals of Statistical Inference II	(Spring '21, '22)
Guest lecture for NCSU Libraries Statistics Power Half Hour	(Spring '22)
Guest lecture for Summer Institute in Biostatistics	(Summer '22)

Duke University, Department of Statistics

Guest lecture for STA561: Probability for Machine Learning	(Spring '22)
--	--------------

University of Michigan, School of Public Health

Taught two lectures for the 2019 Big Data Summer Institute (BDSI); recordings available online at the U-M BDSI 2019 Wiki (https://sph.umich.edu/bdsi/).	('19)
Graduate Student Instructor for BIOSTAT 503: Introduction to Biostatistics.	('14)

ADVISING & MENTORING

PhD students (co-)advised

- Cole Manschot (PhD co-advisor with Eric Laber) (NCSU, expected graduation Spring 2023)
- Jimmy Hickey (PhD co-advisor with Jonathan Williams) (NCSU, expected graduation Spring 2024)
- Joe Zhao (PhD co-advisor with Shu Yang) (NCSU, expected graduation Spring 2023)
- Hyoshin Kim (PhD co-advisor with Sujit Ghosh) (NCSU, expected graduation Spring 2024)

PhD Student committees

- Mohamed Abdelkader Abba (NCSU, expected graduation Spring 2024)
- Sanghyun Choo (NCSU, expected graduation Spring 2023)

Masters students advised

- Annabel Settle (NCSU, expected graduation Spring 2023)

Undergraduate students mentored

- Livia Poppa (NCSU)
- Vrishank Ghosh (NCSU)

PRESENTATIONS

Oral (*upcoming)

1. *Information sharing for efficient inference from different data sources (invited). ENAR Spring Meeting (Mar '23). Nashville, Tennessee.
2. *Distributed inference for spatial extremes modeling in high dimensions (invited). '22 IMS International Conference on Statistics and Data Science (Dec '22). Florence, Italy.
3. *Functional regression with wearable device data: a new lens through data partitioning (invited). Seminar of Columbia University's Department of Biostatistics (Dec '22).
4. *Distributed inference for spatial extremes modeling in high dimensions (invited). KAUST Statistics Workshop (Nov '22). King Abdullah University of Science and Technology, Saudi Arabia.
5. *Mean structure learning with high-dimensional correlated data (invited). AISC (Oct '22). Greensboro, NC.
6. Distributed inference for extreme value analysis of large spatial datasets (invited). COMPSTAT '22 (Aug '22). Bologna, Italy. Participating virtually.
7. Streaming inference with intensively measured longitudinal outcomes from wearable devices (invited). Joint Statistical Meetings (JSM) (Aug '22). Washington, DC.
8. Functional regression with wearable device data (invited). ICSA Canada Chapter Symposium (Jul '22). Banff, Canada.
9. Joint integrative analysis of dependent data sources (invited). ICSA Applied Statistics Symposium (June '22). Gainesville, Florida.
10. Statistical inference for streamed longitudinal data (invited). EcoSta (June '22). Kyoto, Japan. Held virtually.
11. Functional regression with wearable device data: a new lens through data partitioning (invited). Seminar of the University of California – Berkeley Division of Biostatistics (Apr '22). Berkeley, California.

12. Learning mean homogeneity structure in image-on-scalar regression. ENAR Spring Meeting (Mar '22). Houston, Texas.
13. Sleep classification with artificial synthetic imaging using deep learning CNN (invited). AAAS Annual Meeting Symposium (Feb '22). Held virtually.
14. A framework for data integration with dependence and heterogeneity (invited). CMStatistics (Dec '21). London, UK. Held virtually.
15. Data integration meets divide-and-conquer: dealing with heterogeneity and dependence in big data (invited). Seminar of the University of Pittsburgh Department of Statistics (Oct '21). Held virtually.
16. Fused mean structure learning in data integration with dependence. Joint Statistical Meetings (JSM) (Aug '21). Held virtually.
17. Integrated fused mean structure learning with application to image-on-scalar regression (invited). Statistical Methods in Imaging Conference (May '21). Held virtually.
18. Joint integrative analysis of multiple data sources with correlated vector outcomes (invited). ENAR Spring Meeting (Mar '21). Held virtually.
19. Parallel-and-Stream accelerator for regression analysis with big data (invited). Seminar of the North Carolina State University Department of Statistics (Feb '21). Held virtually.
20. Joint integrative analysis of multiple data sources with correlated vector outcomes (invited). Joint seminar of the HEC Département de sciences de la décision and the McGill University Department of Statistics (Feb '21). Held virtually.
21. Accelerated distributed inference through a unified Fiducial and Frequentist paradigm (invited). BFF6.5 Virtual Workshop on Bayesian, Fiducial and Frequentist Statistical Inference (Feb '21). Held virtually.
22. Distributed inference with correlated outcomes (invited). Seminar of the North Carolina State University Department of Statistics (Nov '19). Raleigh, North Carolina.
23. A unifying framework for distributed and integrated inference with high-dimensional correlated outcomes. Workshop on BFF (Bayes, Fiducial and Frequentist) paradigm in data integration, machine learning and applications (Nov '19). Ann Arbor, Michigan.
24. Doubly distributed and integrated inference for correlated data with heterogeneous parameters. Joint Statistical Meetings (JSM) (Jul '19). Denver, Colorado.
25. A distributed and integrated method of moments for high-dimensional correlated data analysis (invited). Workshop on Recent Developments in Statistical Theory and Methods Based on Distributed Computing (May '18). Casa Matemática Oaxaca (CMO), Banff International Research Station for Mathematical Innovation and

Discovery. Oaxaca, Mexico. Recording available at

<http://www.birs.ca/events/2018/5-day-workshops/18w5089/videos>.

26. A distributed and integrated method of moments for high-dimensional correlated data analysis. ENAR Spring Meeting (Mar '18). John Van Ryzin Award winner. Atlanta, Georgia.
27. A distributed and integrated method of moments for high-dimensional correlated data analysis. University of Michigan Data Science Research Forum ('17). Michigan Institute for Data Science (MIDAS). Ann Arbor, Michigan. Recording available at <https://www.youtube.com/watch?v=SGPPfX2T4dI>.

Poster

1. Distributed quadratic inference functions. Joint Statistical Meetings (JSM) ('20). Held virtually.
2. A doubly distributed and integrated method of moments for high-dimensional correlated data analysis. Joint Statistical Meetings (JSM) ('18). Vancouver, Canada.
3. Regression analysis for high-dimensional correlated outcomes. Spatial Statistics Conference ('17). University of Lancaster. Lancaster, England.
4. Change in variance of IGF2 gene methylation associated with three metabolites. ENAR Spring Meeting ('16). Austin, Texas.

Other

Panel Discussant. Fifth Bayesian, Fiducial and Frequentist (BFF5) Conference: Foundations of Data Science ('18). Ann Arbor, Michigan.

PROFESSIONAL SERVICE

American Statistical Association

- IMS contributed session chair, Joint Statistical Meetings (2024).
- Co-organizer and chair of Topic-Contributed Session at the Joint Statistical Meetings: “*Recent Advances in High-Dimensional Estimation and Inference Methods*” (2022)
- Organizer and Chair of Topic-Contributed Session at the Joint Statistical Meetings: “*Integrative inference with data from multiple sources: challenges and new developments*” (2021)
- ASA Section on Statistical Learning and Data Science 2021 Student Paper Award judge (2021)

International Biometric Society

- Chair of Contributed Papers Session at ENAR: “*Clustered Data Methods*” (2022)

Referee

Annals of Applied Statistics, Biometrics, Biostatistics, Computational Statistics and Data Analysis, Journal of the American Statistical Association, Journal of Applied Statistics, Journal of Computational and Graphical Statistics, Journal of Data Science, Journal of Statistical Computation and Simulation, Proceedings of the National Academy of Sciences of the United States of America, Science Advances, Statistics in Medicine, Statistical Methods in Medical Research

DEPARTMENTAL SERVICE

North Carolina State University

1. Seminar Committee, Member (2021), Chair (2022).
2. Departmental Climate Committee, Member (2021-).
3. Departmental Qualifying Exam Committee, Member (2022).
4. Beach Trip Committee, Member (2022).
5. Professional Strategies Working Group, Member (2020-).

University of Michigan, School of Public Health

1. Faculty Search Committee, Student representative (2019-2020)
2. Biostatistics Computing, Social Media and Website Reform Committee, Student representative (2018-2019)
3. Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS), Lead member, Planning Committee (2017-2018)
4. Fifth Bayesian, Fiducial, and Frequentist (BFF5) Conference, Member, Planning Committee (2017-2018)
5. Biostatistics Brown Bag Seminar, Member (2014-2018), President (2016-2018)
6. STATCOM (Statistics in the Community at Michigan) (2015-2017)