

## Emily C. Hector

North Carolina State University

Email: [ehector@ncsu.edu](mailto:ehector@ncsu.edu)

Website: [emilyhector.com](http://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, Environmental health science, Epigenetics, Metabolomics.

### **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. *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)
2. *Excellence in Research Award*, awarded annually to one student in recognition of research excellence. Department of Biostatistics, University of Michigan (2019)
3. *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)
4. *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)
5. *Rackham Conference Travel Grant*, University of Michigan (2016, 2017, 2018, 2019)
6. *Outstanding First-Year Masters Student*, Department of Biostatistics, University of Michigan (2015)
7. *First Class Honors*, McGill University (2014)
8. *Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award (USRA)*, McGill University (2013)
9. *J W McConnell Scholarship* (major award), McGill University (2011-2013)

## PUBLICATIONS

---

### *Peer-reviewed journal articles – Statistical Methodology*

1. **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.
2. **Hector EC**, Song P XK. A distributed and integrated method of moments for high-dimensional correlated data analysis. *Journal of the American Statistical Association* (2020). doi: 10.1080/01621459.2020.1736082.

### *Peer-reviewed journal articles – Statistical Applications*

1. 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.

2. Jansen E, **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.
3. 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.

*Journal articles in review or revision*

1. **Hector EC**, Song P XK (2020+). Joint integrative analysis of multiple data sources with correlated vector outcomes. arXiv, arXiv:2011.14996.
2. **Hector EC\***, Luo L\* and Song P XK (2020+). Parallel-and-stream accelerator for computationally fast supervised learning.
3. **Hector EC** (2020+). Fused mean structure learning in data integration with dependence.
4. Tagelsir Ahmed A, **Hector EC**, Urena-Cirett JL, Mercado-Garcia A, Cantoral A, Hu H, Peterson KE, Téllez-Rojo MM, Martinez-Mier EA (2020+). Early Lead Exposure is Associated with Molar Incisor Hypomineralization (MIH): Results from The Early Life Exposure in Mexico to Environmental Toxicants study.
5. Shi L, Wank M, Chen Y, Wang Y, **Hector EC**, Song P XK (2021+). Sleep Classification with Artificial Synthetic Imaging Data from Empatica E4 Wristband by Convolutional Neural Networks.

*Book chapters*

1. **Hector EC**, Tang L, Zhou L, Song P XK (2021+). Data integration and fusion in the Bayesian, Fiducial and Frequentist framework. Chapter in “Handbook on Bayesian, Fiducial and Frequentist Inference”. Submitted.

\* denotes co-first authorship

## **SOFTWARE**

---

*R packages*

1. DIQIF: Performs joint integrative regression analysis of multiple data sources with correlated vector outcomes. Regression parameters from each data source are estimated using quadratic inference functions, and parameters are selectively

combined following a pre-specified partition using a meta-estimator similar in spirit to Hansen's generalized method of moments.

2. DDIMM: Performs doubly distributed and integrated method of moments regression for high-dimensional correlated responses. Outcomes and subjects are divided into blocks that are analysed using composite likelihood or generalized estimating equations. Block estimates are combined using a meta-estimator similar in spirit to Hansen's generalized method of moments.
3. DIMM: Performs distributed and integrated method of moments regression for high-dimensional correlated responses. Divides outcomes into blocks, analyses blocks using composite likelihood, and combines estimators using a one-step update or an optimal generalized method of moments (GMM).

## **TEACHING**

---

### **North Carolina State University**

ST502: Fundamentals of Statistical Inference II (Spring 2021)

### **University of Michigan, School of Public Health**

*Graduate Student Instructor*

Responsible for preparing and teaching two lectures (Linear Regression and Logistic Regression) for the 2019 Big Data Summer Institute (BDSI); recordings available online at the U-M BDSI 2019 Wiki accessible from <https://sph.umich.edu/bdsi/>. (2019)

BIOSTAT 503: Introduction to Biostatistics. Supervisor: Thomas Braun, PhD (2014)

## **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)

### **Undergraduate students mentored**

- Livia Poppa (NCSU, expected graduation Spring 2023)

## **PRESENTATIONS**

---

### *Oral* (\*upcoming)

1. \*A flexible and computationally fast framework for spatial extremes modelling (invited). International Chinese Statistical Association Applied Statistics Symposium (September 2021). Held virtually.
2. \*Fused mean structure learning in data integration with dependence. Joint Statistical Meetings (JSM) (August 2021). Held virtually.
3. Integrated fused mean structure learning with application to image-on-scalar regression (invited). Statistical Methods in Imaging Conference (May 2021). Held virtually.
4. Joint integrative analysis of multiple data sources with correlated vector outcomes (invited). ENAR Spring Meeting (March 2021). Held virtually.
5. Parallel-and-Stream accelerator for regression analysis with big data (invited). Department of Statistics Seminar (February 2021). North Carolina State University. Held virtually.
6. 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 Statistics Department (February 2021). Held virtually.
7. Accelerated distributed inference through a unified Fiducial and Frequentist paradigm (invited). BFF6.5 Virtual Workshop on Bayesian, Fiducial and Frequentist Statistical Inference (February 2021). Held virtually.
8. Distributed inference with correlated outcomes (invited). Department of Statistics Seminar (November 2019). North Carolina State University. Raleigh, North Carolina.
9. 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 (November 2019). Ann Arbor, Michigan.
10. Doubly distributed and integrated inference for correlated data with heterogeneous parameters. Joint Statistical Meetings (JSM) (July 2019). Denver, Colorado.

11. 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 2018). 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>.
12. A distributed and integrated method of moments for high-dimensional correlated data analysis. ENAR Spring Meeting (March 2018). John Van Ryzin Award winner. Atlanta, Georgia.
13. A distributed and integrated method of moments for high-dimensional correlated data analysis. University of Michigan Data Science Research Forum (2017). 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) (2020). Held virtually.
2. A doubly distributed and integrated method of moments for high-dimensional correlated data analysis. Joint Statistical Meetings (JSM) (2018). Vancouver, Canada.
3. Regression analysis for high-dimensional correlated outcomes. Spatial Statistics Conference (2017). University of Lancaster. Lancaster, England.
4. Change in variance of IGF2 gene methylation associated with three metabolites. ENAR Spring Meeting (2016). Austin, Texas.

#### *Other*

Panel Discussant. Fifth Bayesian, Fiducial and Frequentist (BFF5) Conference: Foundations of Data Science (2018). Ann Arbor, Michigan.

## **PROFESSIONAL SERVICE**

---

### **American Statistical Association**

- ASA Section on Statistical Learning and Data Science 2021 Student Paper Award judge (2021)

### **Sessions Organized**

- Topic-Contributed Session at Joint Statistical Meetings: “*Integrative inference with data from multiple sources: challenges and new developments*” (2021)

## **Referee**

- Referee for Computational Statistics and Data Analysis, Journal of the American Statistical Association, Journal of Computational and Graphical Statistics, Proceedings of the National Academy of Sciences of the United States of America, Statistical Methods in Medical Research, Statistics in Medicine

## **DEPARTMENTAL SERVICE**

---

### **North Carolina State University**

1. Seminar Committee, Member (Spring 2021).
2. Climate Committee, Member (2021-).
3. 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)