BORA JIN

+1 919 448 4611 bora.jin@duke.edu

Department of Statistical Science, Duke University, Durham, NC 27708

RESEARCH INTERESTS

Environmental health, spatial statistics, multivariate data, hierarchical models, latent variables, Bayesian methods

EDUCATION

Ph.D. Candidate in Statistical Science **2018 – 2022 (expected)**
Duke University Durham, NC, USA
Advisors: Amy H. Herring, David Dunson

Master of Applied Statistics **2015 – 2017**
Korea University, 4.25 Grade Point Average (4.5 Scale) Seoul, SOUTH KOREA

Bachelor of Economics in Statistics **2011 – 2015**
Korea University, 4.3 Grade Point Average (4.5 Scale) Seoul, SOUTH KOREA

RESEARCH EXPERIENCE

Bag of DAGs: Flexible & Scalable Modelling of Spatiotemporal Dependence **Present**
with Peruzzi, M., Johndrow, J.E., and Dunson, D.B.

- Propose a computationally efficient approach to construct a well-defined spatial Gaussian process (GP) with the nonstationary covariance using multiple yet simple directed acyclic graphs (DAGs), which leads to computational efficiency, flexibility, and interpretability in point-referenced geostatistical models.
- Develop "bag of DAGs," each of which is chosen to represent a different possible dependence structure, to induce nonstationarity.
- Analyze spatiotemporal variability of fine particulate matter (PM2.5) in California in which a DAG represents a prevailing wind direction causing some associated covariance in the pollutants.
- Study associations between air pollution and hospital visits related to COVID-19.

Scalable Gaussian Processes on Physically Constrained Domains **Present**
with Herring, A.H. and David B.D.

- Motivated by groundwater contamination in which pollutant measurements are collected and meaningful only in a constrained domain, i.e., groundwater bodies with intrinsic geometry.
- Develop the Barrier Overlap-Removal Acyclic Directed Graph GP (BORA-GP), a scalable GP method that incorporates the constrained domain via sparsity-inducing DAGs.
- Enable characterization of dependence in constrained domains by removing an edge in a DAG if a linear path between two points overlaps physical barriers.
- Analyze water pollutant measurements in California collected through the Groundwater Ambient Monitoring and Assessment Program (GAMA).

Bayesian Matrix Completion for Hypothesis Testing **2019 – 2020**
with Dunson, D., Rager, J.E., Reif, D., Engel, S.M., Herring, A.H.

- Adapt Bayesian heteroscedastic nonparametric regression to a multiple hypothesis testing framework.
- Impose a generalized latent factor model to form a non-exchangeable prior for testing.
- Develop a matrix completion method for a latent matrix.
- Tackle sparsity of the ToxCast data using hierarchical framework.
- Enable prediction for non-tested chemical's activity.
- Broaden the definition of activity including heteroscedasticity.

Master's thesis on Bayesian Methods **March 2015 – February 2017**
Korea University **Seoul, SOUTH KOREA**

- Applied Bayesian inference for a seemingly unrelated regression model and examined novel statistical methods on an extended instrumental variables model with random effects using the MCMC method.
- Employed the extension to a nonparametric model using cosine basis functions and the Dirichlet process location-scale mixture for a great deal of flexibility on the proposed model.
- Conducted a real data application explaining Annex I Parties variations in compliance to the Kyoto Protocol.

Prediction of Carbon Emissions in Industrial Setting **June 2016 – August 2016**
Korean Environment Ministry **Seoul, SOUTH KOREA**

- Trained toward a generalist dealing with global environmental issues in both national and international settings through the 8th International Environmental Expert Training Program.
- Applied statistical prediction analysis in studying afforestation practices within industrial complex areas.
- Developed strategic programs to implement carbon reduction targets of industrial complexes.

Clustering of Cancer Patients' Symptoms **March 2016 – August 2016**
College of Nursing at Chungnam National University **Seoul, SOUTH KOREA**

- Conducted clustering analysis to see if symptoms are divided into particular clusters as expected in a theory.
- Measured the effect of symptom clusters on the quality of life through physical functions.
- Provided expertise on all aspects from the interpretation of statistical results to the presentation of statistics and graphics.

Emotionality of Language in Online Platforms **October 2014 – March 2015**
MezzoMedia & SungKyunKwan University **Seoul, SOUTH KOREA**

- Contributed to the development of a Korean morphological analyser.
- Embedded factor analysis and principal component analysis to determine appropriate weights of frequently used morphemes based on their verbal and social context.
- Designed and interpreted quantitative research examining the emotionality of language in online platforms.

WORK EXPERIENCE

ENVIRONMENTAL STATISTICS

Internship in Chemicals and Waste **February 2017 – August 2017**
UN Environment **Geneva, SWITZERLAND**

- Managed national reports that Parties are obliged to submit under Basel and Stockholm Conventions and analysed national reporting data so as to identify regional patterns and temporal trends.
- Designed a query system to facilitate proper visualizations and the use of collected data from national reports.
- Participated in the formulation of indicators with regards to Sustainable Development Goals.

BIG DATA & FINANCE

Internship in the Division of Budget and Finance **January 2018 – July 2018**
International Atomic Energy Agency **Vienna, AUSTRIA**

- Managed data compliance of requests for all types of procurement and payment in the Agency.
- Analysed collected data as a member of Master Data Management Team.
- Assisted a clean-up project of the Agency's bank and branch pages through Oracle sql and MS Access.
- Published monthly infographics on activities of the whole division.

Internship in Banking and Finance, Market Surveillance Department **June 2014 – July 2014**
Korea Exchange (KRX) **Seoul, SOUTH KOREA**

- Focused upon the detection of unfair transactions, particularly with regards to high turnover volume accounts.
- Conducted data management including updating, arranging and organising big data.
- Participated in developing case-specific restrictions and market-wide regulatory practices.

TEACHING EXPERIENCE

Head Teaching Assistant **Present**
Duke University **Durham, NC, USA**

- Theory and Methods of Statistical Learning and Inference (STA432)

Teaching Assistant **January 2019 – May 2019**
Duke University **Durham, NC, USA**

- Statistics (STA250)

Teaching Assistant **September 2015 – December 2015**
Korea University **Seoul, SOUTH KOREA**

- Introduction to Probability Theory (STAT201)
- Topics in Mathematical Statistics (STAT412)

Mentoring in Mathematics **January 2013**
Samsung **Seoul, SOUTH KOREA**

KEY SKILLS

- R, MATLAB, Python, SAS, MySQL, MS Office, LATEX
- English (Professional Proficiency), Korean (Native)

HONOURS

NC ASA Student Travel Awards 2021
Global Korea Scholarship 2018 – 2020
First Prize in the Graduate Paper Section at the Korean Statistical Society's Annual Conference 2016
Yangcheon Foundation Scholarship for Academic Excellence 2016
So-Mang Presbyterian Church Scholarship for Academic Merit 2016
Second Prize in the Graduate Poster Section at the Korean Statistical Society's Annual Conference 2015
The Dean's Award for Academic Merit 2011 – 2014
Ministry of Gender Equality and Family Affairs Minister's Honour Award 2012
Seoul National University President's Prize 2011

PUBLICATIONS

Jin, B., Dunson, D., Rager, J.E., Reif, D., Engel, S.M., Herring, A.H. (2020+). Bayesian Matrix Completion for Hypothesis Testing. *submitted* ArXiv

WORKING PAPERS

Jin, B., Peruzzi, M., Johndrow, J.E., Dunson, D.B. Bag of DAGs: Flexible & Scalable Modelling of Spatiotemporal Dependence

Jin, B., Herring, A.H., Dunson, D.B. Scalable Gaussian Processes on Physically Constrained Domains