Wayne Yu Wang

wayneyw@umich.edu 734.277.5499 ywa136.github.io

Education

- **University of Michigan** — Ann Arbor, MI
  Ph.D. in Statistics — *Sep '18 - Aug '22 (expected)*
  Dissertation: Statistical and geometric models for complex spatio-temporal data
  Advisors: Drs. Alfred Hero and Yang Chen
- **University of British Columbia** — Vancouver, Canada
  M.S. in Statistics — *Sep '16 - Aug '18*
  Advisors: Drs. Jim Zidek and Nhu Le
- **Simon Fraser University** — Vancouver, Canada
  B.S. in Actuarial Science — *Sep '12 - Sep '16*

Technical Skills

- **Scientific Programming**: Julia, Python, MATLAB.
- **Statistical Software**: R (tidyverse, dplyr, ggplot2, Rcpp, STAN/RStan).
- **Machine Learning Frameworks**: TensorFlow, Keras, scikit-learn, Flux.
- **Tools**: AWS, Git/GitHub, SQL (MySQL, T-SQL), Hadoop (PySpark, MapReduce), MPI.

Research Experience

- **Department of Statistics and EECS, University of Michigan**, Graduate Researcher — *Aug '18 - present*
  - **High-dimensional Gaussian Graphical Models for Tensor-Variate Data**: Proposed a novel physics-driven generative model for high-dimensional multiway/tensor-variate data, leading to sparse Kronecker structures on the inverse covariance matrix. Designed efficient optimization algorithms for learning the underlying inverse covariance matrix. *Tools: Julia, R* (☞Papers published at AISTATS '20 [6], ICML '21 [3], and NeurIPS '21 [1]; Developed an associated Julia package called `TensorGraphicalModels`).
  - **Time-Varying Topic Models**: Developed a framework for topic modeling of time-varying corpora, combining statistical models with computational geometric methods. Applied the framework to Twitter data for understanding public opinions on COVID-19. Currently extending the methods to supervised learning settings and the application to text data from TalkLife, a peer support social network for mental health. *Tools: Python, R, Spark, Hadoop, SQL* (☞Paper published at Harvard Data Science Review [4]; Developed an online exploratory analysis/visualization tool using `R Shiny`).
  - **Bayesian Point Processes for Solar Flare Forecasting**: Proposed a hierarchical Bayesian model based on marked Hawkes processes for forecasting the onset of solar flares. *Tools: R, STAN.*
- **Statistical Sciences Group, Los Alamos National Laboratory**, Research Intern — *May '21 - Aug '21*
  - **Streaming Distributed PCA for Exascale Climate and Space Sciences**: Designed an efficient streaming & distributed PCA algorithm for in situ scientific discovery (online data analysis and visualization while the scientific simulation is running). Applied the algorithm to exascale data generated from climate (E3SM) and space weather (SHIELDS/SWMF) simulations. *Tools: Julia, MapReduce, MPI* (☞Paper published at SC '21 [2]; Developed an associated Julia package called `TributaryPCA`).
- **Department of Statistics, University of British Columbia**, Graduate Researcher — *Sep '16 - Aug '18*
  - **Determinantal Point Processes for Spatial Design**: Proposed a randomized design strategy for spatial monitoring networks via determinantal point processes (DPP). Studied the relationship between DPP approximations and combinatorial optimization problems, where the approximation efficiency is established via record value/extreme value theory. *Tools: R* (☞Papers published at several applied statistical journals including [5]).
- **Department of Statistics and Biomedical Engineering, Simon Fraser University**, Undergraduate Researcher — *May '15 - Sep '15*
  - **Statistical Imputation of SNP Information for Alzheimer's Disease**: Implemented a pipeline for imputation of un-genotyped SNPs from the Alzheimer's Disease Neuroimaging Initiative data, based on variants of hidden Markov models (HMMs). *Tools: R* (☞Undergraduate Student Research Award from Vice-President Research, SFU).

Selected Publications (Full list at google scholar)

[1] **Y. Wang**, A. O. Hero, Multiway Ensemble Kalman Filter, *NeurIPS '21: Workshop on Machine Learning and the Physical Sciences.*

[2] **Y. Wang**, N. Klein, S. Morley, V. Jordanova, M. Henderson, A. Biswas, E. Lawrence, TributaryPCA: Distributed, Streaming PCA for In Situ Dimension Reduction with Application to Space Weather Simulations, *SC '21: International Workshop on Data Analysis and Reduction for Big Scientific Data.*

[3] **Y. Wang**, A. O. Hero, SG-PALM: a Fast Physically Interpretable Tensor Graphical Model, *ICML '21*, [arxiv].

[4] **Y. Wang**, C. Hougen, B. Oselio, W. Dempsey, A. O. Hero, A Geometry-Driven Longitudinal Topic Model, *Harvard Data Science Review*, 2021, [journal web].

[5] **Y. Wang**, J. V. Zidek, N. D. Le, Approximately Optimal Subset Selection for Statistical Design and Modelling, *Journal of Statistical Computation and Simulation*, 2021, [journal web].

[6] **Y. Wang**, B. Jang, A. O. Hero, The Sylvester Graphical Lasso (SyGlasso), *AISTATS '20*, [arxiv]