

SUMMARY

Data scientist and applied researcher with a PhD in Statistics and 6+ years of experience in statistics and machine learning. A wide spectrum of expertise in experimental design, causal inference, high-dimensional/spatio-temporal data analysis, optimization. Proven ability to apply analytical methods to solve practical product problems with measurable impact, develop innovative data-driven solutions at scale, and communicate complex technical concepts to a variety of audiences (e.g., Product Managers, Engineers, Researchers, etc) in cross-functional projects.

TECHNICAL SKILLS

- **Scientific Programming:** Julia, Python, MATLAB.
- **Statistical Software:** R (tidyverse, dplyr, ggplot2, Rcpp, STAN/RStan).
- **Machine Learning Frameworks:** TensorFlow, scikit-learn.
- **Tools:** GCP, Git, Mercurial, SQL (GoogleSQL, BigQuery), Hadoop/Spark, MPI.

PROFESSIONAL EXPERIENCE

- **Google, Data Scientist** Jul '22 - present
 - **Developing Model-based Measurement:** Data Science PoC for YouTube Ads relevance optimization and measurement, developing and productionizing machine learning model-based measurement framework, trained on online surveys and human evals, for measuring and optimizing users' perceived relevance towards ads. *Tools: R, GoogleSQL* (☞The resulting metrics have increased the metric sensitivity in A/B experiments by 2x to 12x, effectively saved the team from having to increase the survey load by at least 4x, which amounts to roughly 53 million ads slot saved at minimum; Developed associated internal R packages for general model-based metric construction).
 - **Optimizing Users' Overall Perception:** Developing methods for optimally mapping users' online survey responses to numerical scores (considered as "ground truth" labels for metrics and models) via Bayesian optimization and latent variable models. *Tools: Python, GoogleSQL* (☞The new survey response labels increased sensitivity of the survey based metrics in A/B experiments by more than 30%).
 - **Designing and Analyzing Human Evals:** Designing human evaluation templates and guidelines. Implementing statistical models for measuring and improving rater agreement and accuracy of the evals. *Tools: Python, R, GoogleSQL* (☞Revamped evaluation templates and measurement techniques have boosted the agreement among human raters by more than 3X, effectively reducing evaluation cost and the noise of the ratings subsequently used in machine learning model training).
- **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 online analysis and visualization of exascale data generated from climate and space weather simulations. *Tools: Julia, MapReduce, MPI* (☞Paper published at ACM/IEEE Supercomputing Conference '21; Developed an associated open-source Julia package called **TributaryPCA**).

RESEARCH EXPERIENCE

6+ years of research experience in statistics, machine learning, and optimization, with broad applications in text & image analysis and space & climate science. 10+ peer-reviewed publications in top statistical journals and machine learning conferences with 60 citations (☞Google Scholar page). Selected research projects include:

- **High-dimensional Gaussian Graphical Models for Tensor-Variate Data:** Proposed a novel physics-driven generative model for high-dimensional multiway/tensor-variate data. Designed efficient optimization algorithms for learning the underlying statistical model. *Tools: Julia, R* (☞Papers published at AISTATS '20, ICML '21, NeurIPS '21, and Statistics Surveys; Developed an open-source 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. *Tools: Python, Spark, Hadoop, SQL* (☞Paper published at Harvard Data Science Review; Developed an online exploratory analysis/visualization tool using **R Shiny**).
- **Deep Learning Ensembles for Solar Flare Forecasting:** Proposed an ensemble method combining LSTM and CNN for classification of flare-imminent active regions that produce strong flares in the succeeding 24hr window. *Tools: Python* (☞Paper published at The Astrophysical Journal).

EDUCATION

- **University of Michigan** Ann Arbor, MI
Ph.D. in Statistics Sep '18 - Jul '22
Dissertation: Interpretable and Scalable Graphical Models for Complex Spatio-temporal Processes
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. with Distinction in Actuarial Science (Completed SOA Exams P, FM, and MFE) Sep '12 - May '16