Hu Sun

husun@umich.edu | +1 734-834-1630 | husun0822.github.io | Hu Sun |

Education

University of Michigan, Ann Arbor — Ann Arbor, U.S.

Ph.D. in Statistics (Advisor: Yang Chen, GPA: 4.0/4.0) — *Aug. 2020 – Jun 2024 (Expected)*

- Research interest: Tensor Data Model; Spatio-Temporal Process; Astro-statistics.

M.S. in Applied Statistics (GPA: 4.0/4.0) — *Aug. 2018 – May 2020*

- Awards: Outstanding First Year Master Student Award

University of Oxford — England, U.K.

Visiting Student in Economics (GPA: First-Class (Oxford Scale)) — *Oct. 2016 – Jun. 2017*

- Awards: $\$15,000$ Oxford Visiting Student Fellowship

Xiamen University — Xiamen, China

B.A. in Economics (GPA: 3.92/4.0, Rank: 1/47) — *Aug. 2014 – Jun. 2018*

- Thesis: *Self-Confidence, Attribution Bias and their Implications on Job Applications*
- Awards: Best Student Thesis Award (2018)

Research Experience (See my full publication list here)

Tensor Completion Method for High-Dimensional Spatio-Temporal Data [website]

Department of Statistics, University of Michigan, Ann Arbor — *Aug. 2020 – Jul. 2022*

- Proposed a novel tensor completion model named *VISTA* for missing value imputation in large spatial-temporal datasets that guarantees spatial and temporal smoothness. A scalable estimation algorithm is developed with both theoretical guarantees and superior performances over existing tensor completion methods in both simulated and real datasets from geophysics.
- Publications:
  - **Sun, H.**, Hua, Z., Ren, J., Zou, S., Sun, Y., & Chen, Y. (2022). Matrix Completion Methods for the Total Electron Content Video Reconstruction. *Annals of Applied Statistics*, 16(3), pp.1333-1358.
  - **Sun, H.**, Chen, Y., Zou, S., Ren, J., Chang, Y., Wang, Z., & Coster, A., (2023). Complete Global Total Electron Content (TEC) Map Dataset based on a Video Imputation Algorithm VISTA. *Scientific Data, in press*.
  - Wang, Z., Zou, S., **Sun, H.**, & Chen, Y, (2023). Forecast Global Ionospheric TEC: Apply Modified U-Net on VISTA TEC Dataset. *Submitted*.

Tensor Auto-regression with Auxiliary Vector Time Series Data

Department of Statistics, University of Michigan, Ann Arbor — *Jun. 2021 – Present*

- Proposed a novel semi-parametric auto-regressive model for matrix-valued time-series data with auxiliary vector covariates. We accompany the matrix predictors with auto-regressive coefficient of Kronecker-product form and the auxiliary covariates with functional parameters from a Reproducing Kernel Hilbert Space. We establish the joint asymptotics of the model estimators in fixed and high dimensional settings and demonstrate the performances with extensive simulation & real data application.
- Publication:
  - **Sun, H.**, Shang, Z., & Chen, Y., (2023). Matrix Auto-regressive Model with Vector Time Series Covariates for Spatio-Temporal Data. *Ready for submission*.

Scalar-on-Tensor Regression with Contracted Tensor Gaussian Process

Department of Statistics, University of Michigan, Ann Arbor — *Apr. 2022 – Jan. 2023*

- Propose a Gaussian Process (GP) Regression model for scalar responses with tensor covariates. The regression model has a tensor contraction step and a multi-linear GP step that first reduces the size of a tensor to a smaller core tensor and then fit the tensor GP on the core tensor for scalable, memory-efficient computation. We propose an alternating proximal gradient descent method with theoretical guarantees and extensively test the method on simulation data and astronomy data for flare forecasting.
- Publication:
  - **Sun, H.**, Manchester, W., Meng, J., Liu, Y., & Chen, Y., (2023). Tensor Gaussian Process with Contraction for Multi-Channel Imaging Analysis. *Submitted for review*.

Machine Learning Prediction for Solar Flare with Spatio-Temporal Data

Department of Statistics, University of Michigan, Ann Arbor *Jan. 2019 – Present*

- Applied Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) model for forecasting an ultra-rare space weather event called solar flare. We've achieved state-of-the-art prediction performances in the current research field of astro-statistics.
- Proposed a novel surrogate model for interpreting the prediction made by the LSTM model and locate the precursors within the time-series data based on an unsupervised learning approach.
- Introduced Spatial Statistics and Topological Data Analysis (TDA) techniques to derive new hand-crafted features from the imaging data. We have achieved comparable performances with other deep learning based model but have greatly improved the model interpretability.
- Papers:
  - **Sun, H.**, Manchester IV, W. B., & Chen, Y. (2021). Improved and Interpretable Solar Flare Predictions With Spatial and Topological Features of the Polarity Inversion Line Masked Magnetograms. *Space Weather*, 19(12), p.e2021SW002837.
  - Jiao, Z., **Sun, H.**, Wang, X., Manchester, W., Gombosi, T., Hero, A., & Chen, Y. (2020). Solar Flare Intensity Prediction with Machine Learning Models. *Space Weather*, 18(7).
  - **Sun, H.**, Manchester, W., Jiao, Z., Wang, X., & Chen, Y. (2019). Interpreting LSTM Prediction on Solar Flare Eruption with Time-series Clustering. *arXiv preprint*.

Contributed/Invited Talks & Posters

- Matrix Auto-regressive Model With Vector Time-series Covariates. *INFORMS Annual Meeting*. Oct, 2022.
- Improved and Interpretable Solar Flare Predictions with Spatial & Topological Features of the Polarity-Inversion-Line Masked Magnetograms. *Department of Computer Science, Georgia State University*. Sept, 2021.
- Improving and Interpreting Flare Prediction with Spatial Statistics Analysis of the Magnet Field Data. *Joint Statistical Meeting (JSM)*. Aug, 2021.
- Video Imputation Model in the Context of Space Weather Monitoring. *Jet Propulsion Laboratory (JPL)*. June, 2021.
- Improved and Interpretable Solar Flare Predictions with Spatial and Topological Features of the Polarity-Inversion-Line Masked Magnetograms. *Conference on Applications of Statistical Methods and Machine Learning in the Space Sciences, Boulder, Colorado*. May, 2021.
- Video Imputation and Prediction Models in Context of Space Weather Monitoring. *CHASC Astro-statistics Seminar Series, Harvard University*. Apr, 2021.
- (Poster) Interpreting LSTM Prediction on Solar Flare Eruption with Time-series Clustering. *American Geophysical Union (AGU) Meeting*. Dec, 2020.

Technical Skills

- Machine Learning: Python (TensorFlow, PyTorch, OOP), C++ (Parallel Computing)
- Data Analysis: R (rSTAN, Rcpp), SQL, MATLAB, STATA
- Language: English (fluent, TOEFL 119/120), Mandarin