Hu Sun

husun@umich.edu | +1 734-834-1630 | husun-leo.com | Hu Sun | Hu Sun
Apt 4C, 125 W. Hoover Ave., Ann Arbor, MI, 48103 |

Education

University of Michigan, Ann Arbor — Ann Arbor, U.S.
Ph.D. in Statistics (Advisor: Yang Chen, GPA: 4.0/4.0) — *Aug. 2020 – May 2024 (Expected)*

- Research interest: Spatial-Temporal Process Modeling, Interpretable Machine Learning, Tensor Data Modeling, Time-series Forecast and Astro-statistics.
- Core courses taken: Linear Models, Statistical Learning, Computation and Optimization in Statistics, Monte Carlo Methods, Statistical Inference, Large Sample Theory, Stochastic Process, Computer Vision, Reinforcement Learning, Parallel Computing
- Courses served as TA: Statistical Learning (Graduate Level), Probability and Distribution Theory (Graduate Level), Computational Methods in Statistics and Data Science, Introduction to Theoretical Statistics.

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

- Awards: Received Outstanding First Year Master Student Award (awarded to top 5 master students in GPA)
- Core courses taken: Statistical Computing, Bayesian Statistics, Nonparametric Statistics

University of Oxford — England, U.K.
Visiting Student in Economics (GPA: First-Class (Oxford Scale)) — *Oct. 2016 – Jun. 2017*

- Studied at the Pembroke College with Oxford Visiting Student Fellowship (awarded to 6 students a year in China)

Xiamen University — Xiamen, China
B.A. in Economics (GPA: 3.92/4.0) — *Aug. 2014 – Jun. 2018*

- Thesis title: *Self-Confidence, Attribution Bias and their Implications on Job Applications*
- Awards: Best Student Paper Award (2018); Two-time (2014, 2015) Dean's List Winner

Research Experience (See my full publication list here)

Tensor Completion Method for High-Dimensional Spatio-Temporal Data
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. We made an interactive algorithm dashboard with *R Shiny* available here.
- Paper:
  - **Sun, H.**, Hua, Z., Ren, J., Zou, S., Sun, Y., & Chen, Y. (2021). 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., (2022). Complete Global Total Electron Content (TEC) Map Dataset based on a Video Imputation Algorithm VISTA. *Under minor revision at Scientific Data.*
  - Zou, S., Ren, J., Wang, Z., **Sun, H.**, & Chen, Y. (2021). Impact of Storm-Enhanced Density (SED) on Ion Upflow Fluxes during Geomagnetic Storm. *Frontiers in Astronomy and Space Sciences*, p.162.

Auto-regressive Model for Matrix-Valued Data with Auxiliary Vector Covariates
Department of Statistics, University of Michigan, Ann Arbor — *Jun. 2021 – Present*

- Proposed a novel auto-regressive model for matrix-valued time-series data (data observed on a matrix grid across time) with auxiliary vector covariates (e.g. global quantities shared by all grid points). We utilize a Kronecker product formulation to greatly reduce the dimensionality and accommodate the auxiliary covariates by estimating a linear mapping function from a Reproducing Kernel Hilbert Space. We achieved decent prediction performance on both simulated and real physics data with great model interpretability.
- Paper:
  - **Sun, H.**, Shang, Z., & Chen, Y., (2022). Matrix Auto-regressive Model with Vector Time-series Covariates. *Working Paper.*

Machine Learning Prediction for Solar Flare with Spatio-Temporal Data

Department of Statistics & CLaSP, 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