Yu Chen

ychen.adastra@gmail.com | 256.585.7439 | LinkedIn/yu-chen-astra | ychenxastra.github.io

EDUCATION

Doctor of Philosophy, Space Science, The Univ. of Alabama in Huntsville — June 2018 - Dec 2020
Master of Science, Space Science, The Univ. of Alabama in Huntsville — Aug 2015 - May 2018
Bachelor of Science, Atmospheric Science, Nanjing Univ. of Info. Sci. and Tech. — Sep 2011 - June 2015

WORK EXPERIENCE

CENTER FOR SPACE PLASMA AND AERONOMIC RESEARCH — Huntsville, AL
Research Scientist — Jan 2023 - present
Postdoctoral Researcher — Jan 2021 - Dec 2022

- **Analyzed** extensive high-resolution spacecraft datasets spanning over 50 years (over 50GB) and **designed** data science **algorithms** with scientific computing applications to investigate solar wind structures. **Applied time-series, statistical analyses, and case study methodologies** to derive comprehensive insights.
- **Visualized** data products and **summarized** findings into **20 peer-reviewed journal publications**. **Presented** key insights through oral presentations and posters at large **conferences** for **18 times**, effectively communicating complex information to both expert and non-expert audiences.
- **Administered and updated** the small-scale magnetic flux rope database at fluxrope.info, ensuring public accessibility. **Developed and managed** the team website, facilitating dissemination of valuable information.
- **Led** two national research grants as **Principal Investigator**, contributed as a **Co-Investigator** to multiple projects over $2 million. **Collaborated** with scholars across institutions and **mentored** a non-physics undergraduate in the National Science Foundation (NSF) funded undergraduate research program.
- Served as a **peer reviewer** for NSF, NASA panels, and the Astrophysical Journal, ensuring the integrity of science through rigorous evaluations and **constructive feedback** for around 20 articles and proposals.

PROJECTS

PYGS: A PYTHON PACKAGE FOR FLUX ROPE ANALYSIS — Python, Numpy, Pandas, Scipy, Matplotlib

- **Optimized** the existing detection algorithm by incorporating new theoretical concepts and complex data analytics. **Migrated** separate Matlab-based techniques to Python 3, refactored and merged over 30 scripts into a **modular Python library**, substantially minimizing manual intervention in data processing workflows.
- **Implemented technical upgrades** on several modules including enhancing the cleaning of raw data, reducing both time and space complexity. **Streamlined the architecture** to boost operational efficiency, resulting in a **100x faster** in performance and significantly broadening the applicability for various analytical needs.
- Released the open-source package on GitHub, successfully met the standards and listed on the Python in Heliophysics Community website. **Provided tutoring** and comprehensive instructions to users through **detailed documentation**, and disseminated the package to the wider scientific community.

DATABASE OF SMALL-SCALE MAGNETIC FLUX ROPE (SFR) — Python, Matlab, Excel

- **Implemented detection** for large spacecraft datasets exceeding 50 GB and **designed** sophisticated filtering algorithms to sift through over 200k event candidates. Aggregated qualified SFRs into an online database.
- Utilized **statistical methods**, such as linear regression, to conduct an **in-depth analysis** of distinctive properties. **Presented** parameters and time-series variations in various visual formats, enhancing data interpretability.
- Expertly interpreted variances among different groups, provided clear, **data-driven insights** for formulating and testing hypotheses. **Summarized findings** through detailed publications and presentations, thereby enriching the scientific community's understanding of the object.

ANALYSES OF FOOTBALL MATCHES AND PREDICTING RESULTS — Python, Scikit-learn, Seaborn

- **Analyzed** the Kaggle dataset of the English Premier League (spanning over 20 years), with a special focus on Manchester United. **Visualized statistics** to uncover various factors influencing team performance.
- **Constructed machine learning pipelines** to predict match outcomes using various models such as Naive Bayes, Adaboost, and Random Forest, etc. **Assessed the optimal model** based on metrics and applicability.
- **Refined model accuracy** through advanced techniques like rolling averages, strategic feature selection, and principal component analysis (PCA). Integrated empirical findings with real-world scenarios and conducted comprehensive analyses to address the underlying causes of suboptimal model performance.

SKILLS

Programming Languages: Python, Matlab, MySQL, C/C++
Technology and Tools: GitHub, LaTeX, Jupyter Notebook, Microsoft Office