

Yu Chen

yuchen.adastra@gmail.com | 256.585.7439 | [LinkedIn/yu-chen-astara](https://www.linkedin.com/in/yu-chen-astara) | yuchenxastara.github.io

EDUCATION

Doctor of Philosophy, Space Science, THE UNIV. OF ALABAMA IN HUNTSVILLE
Master of Science, Space Science, THE UNIV. OF ALABAMA IN HUNTSVILLE
Bachelor of Science, Atmospheric Science, NANJING UNIV. OF INFO. SCI. AND TECH.

June 2018 - Dec 2020
Aug 2015 - May 2018
Sep 2011 - June 2015

WORK EXPERIENCE

CENTER FOR SPACE PLASMA AND AERONOMIC RESEARCH

RESEARCH SCIENTIST

POSTDOCTORAL RESEARCHER

Huntsville, AL

Jan 2023 - present

Jan 2021 - Dec 2022

- Utilized **Python** to process and analyze high-resolution **raw datasets** spanning over 50 years (over 50GB). Designed **data science algorithms** with scientific computing applications to identify and catalog typical events, **establishing** a comprehensive local event **database** containing over 140,000 entries.
- Analyzed these typical events by **extracting** and deriving 60+ event **characteristics** for each entry. Applied **time-series, statistical analyses, and case study** methodologies to derive valuable **data-driven insights**, enhancing community's understanding of the phenomenon and proposing hypotheses for underlying relations.
- Developed an **open-source Python package** for studying typical events, enabling **automated output** of characteristics and figures, significantly reducing users' workload by minimizing redundant manual inputs.
- Employed **MATLAB and Matplotlib** for data product **visualization**, summarizing findings into **20 peer-reviewed journal publications**. Presented key insights at large **conferences** for **18 times**, effectively communicating complex information to both expert and non-expert audiences.
- Led **two national research grants** as Principal Investigator, collaborated on projects exceeding \$2 million, and **mentored** underrepresented **students** in the NSF-funded research program, contributing to both research advancements and educational growth.

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 ↗

PYTHON, MATLAB, EXCEL, POSTGRESQL, HTML

- Utilized **Python** to scratch event from large spacecraft datasets exceeding 50 GB. **Designed** sophisticated **filtering algorithms** to sift through over 200k candidates, laying the foundation for the local database.
- Managed the database using **PostgreSQL**, aggregating selected events in an **online platform**. **Maintained and updated** the online database via **HTML** to facilitate efficient and convenient user inquiries.
- Delivered **product properties** via in-depth analysis like linear regression. Provided common parameters in multiple formats to **assist users** in understanding and extracting information relevant to their topics of interest.

ANALYSES OF FOOTBALL MATCHES AND PREDICTING RESULTS ↗

PYTHON, PANDAS, 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 further investigations to address the underlying causes of suboptimal model performance.

SKILLS

Programming Languages: Python, MATLAB, SQL, C/C++

Technology and Tools: Git/GitHub, Linux/Unix, LaTeX, Jupyter Notebook, Microsoft Office