

256.585.7439
ychen.adastra@gmail.com

Yu Chen
New York, NY

ychenxastra.github.io
LinkedIn: yu-chen-astra

EDUCATION

Doctor of Philosophy, Space Science, The University of Alabama in Huntsville
Master of Science, Space Science, The University of Alabama in Huntsville
Bachelor of Science, Atmospheric Science, Nanjing Univ. of Info. Sci. & Tech.

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

EXPERIENCE

Center for Space Plasma and Aeronomic Research

Huntsville, AL

Research Scientist

Jan 2023 — present

Postdoctoral Researcher

Jan 2021 — Dec 2022

- Processed and analyzed over 50GB of high-resolution raw data using Python with specially designed algorithms to identify typical events, improving efficiency and accuracy by 40%.
- Developed an open-source Python package for event studies, automating meaningful information extraction and characteristic visualization, reducing manual input and user workload from months to minutes.
- Established a local database over 140,000 entries and managed it using PostgreSQL and PySpark, enabling efficient querying and data transformation and reducing manual data handling by 30%.
- Evaluated stakeholders needs and extracted 60+ characteristics per entry, applied time-series and statistical analyses to reveal patterns and relationships, deriving insights for strategic decision-making processes.
- Employed MATLAB, Matplotlib, and Tableau for data visualization, aiding in understanding metrics and trends.
- Presented complex data findings in a clear and accessible manner, culminating in 20 articles in peer-reviewed journals and 18 conference presentations, effectively communicating with non-technical stakeholders.
- Led two national research grants as Principal Investigator and collaborated on multi-million-dollar projects, mentored students in an NSF-funded program, contributing to research and educational development.

PROJECTS

PyGS: a Python package for typical event analysis

Python, Numpy, Pandas, Scipy, Matplotlib

- Optimized analyzing models with advanced analytics, migrated separate Matlab-based techniques, refactored and merged 30+ scripts into a Python library, reducing manual data processing and script execution efforts in event identification.
- Upgraded ETL pipelines and streamlined system architecture, significantly reducing time and space complexity, leading to a 100x performance increase and broadened analytical applicability.
- Released the Python package on GitHub, ensuring adherence to Heliophysics community standards, provided detailed documentation and tutorials, and promoted its use in the scientific community.

Database of small-scale magnetic flux rope

Python, Matlab, Excel, PostgreSQL, HTML

- Employed Python to process large datasets spanning over 50 years to pinpoint typical events. Developed sophisticated filtering algorithms to process over 200k candidates, contributing to the establishment of a robust local database.
- Aggregating selected events from the local database to an online platform. Maintained and updated the online database via HTML and Excel to facilitate efficient and convenient user inquiries.
- Conducted in-depth analyses such as linear regression, to characterize product properties. Provided parameters in various formats, assisting users in understanding and extracting relevant information for their specific needs.

Analyses of Football Matches and Predicting Results

Python, Pandas, Scikit-learn, Seaborn

- Analyzed 20+ years of English Premier League data, focusing on Manchester United, and visualized statistics to identify key and controversial factors impacting team performance.
- Developed machine learning pipelines with models like Naive Bayes and Random Forest to predict match outcomes, selecting the best model based on performance metrics and applicability.
- Enhanced model accuracy using rolling averages, feature selection, and PCA, increasing predictive accuracy by 12%. Integrated findings with real-world scenarios to explore causes of suboptimal performance and provide actionable insights.

TECHNICAL SKILLS

Programming Python (Numpy, Pandas, Scipy, Matplotlib, Scikit-Learn, Seaborn, PySpark), MATLAB, C/C++
Tools & Software SQL (MySQL & PostgreSQL), Tableau, Latex, Git, Jupyter Notebook, Excel
Platforms & OS Linux, MacOS, GitHub, AWS