Satellite Data Analysis: TESS, ALMA, HAWKI and JWST

Alexandre Baptista
IT Department
Faculty of Sciences of the University of Lisbon
Portugal
fc64506@alunos.ciencias.ulisboa.pt

Gonçalo Moreno
IT Department
Faculty of Sciences of the University of Lisbon
Portugal
fc49783@alunos.ciencias.ulisboa.pt

Helton Mendonça
IT Department
Faculty of Sciences of the University of Lisbon
Portugal
fc56870@alunos.ciencias.ulisboa.pt

Marcos Torres
IT Department
Faculty of Sciences of the University of Lisbon
Portugal
fc62746@alunos.ciencias.ulisboa.pt

ABSTRACT

Advances in multi-wavelength astronomical observations have revolutionized our understanding of astrophysical phenomena by combining data from space- and ground-based telescopes. This study presents a systematic analysis of observations from the Transiting Exoplanet Survey Satellite (TESS), the Atacama Large Millimeter/submillimeter Array (ALMA), the High Acuity Wide-field K-band Imager (HAWK-I), and the James Webb Space Telescope (JWST), leveraging their complementary capabilities to explore stellar, exoplanetary, and extragalactic systems.

We develop a robust Python-based pipeline utilizing Astroquery for data retrieval, Astropy for coordinate transformations and unit management, Pandas for structured data processing, and NumPy for numerical computations.

As a result of cross-matching observations across all four facilities (TESS, ALMA, HAWK-I, and JWST) and applying coordinate-based alignment criteria to target overlapping fields, we identified 10 high-probability coincident observation spots.

CCS CONCEPTS

• Computing methodologies → Massive data analysis; Data mining; Scientific visualization.

• Applied computing → Astronomy; Physics; Data analysis tools.

• Information systems → Data extraction and integration; Digital libraries and archives.

• Software and its engineering → Software libraries and repositories; Scripting languages (Python).

KEYWORDS

Astroquery, TESS, ALMA, JWST, HAWK-I, Data analysis, Astropy.

ACM Reference format:

Alexandre Baptista, Gonçalo Moreno, Helton Mendonça, and Marcos Torres. 2025. Satellite Data Analysis: TESS, ALMA, HAWKI, and JWST. https://github.com/GoncaloMoreno/IPAI_Project

1. Introduction

The exploration of our universe has entered a revolutionary phase with the advent of advanced space telescopes and ground-based observatories. Modern astronomy now routinely combines observations across the electromagnetic spectrum, from optical and infrared to millimeter wavelengths, to unravel the mysteries of celestial objects [1]. This multi-wavelength approach has become fundamental to nearly all areas of astrophysical research, from studying planetary systems in our galactic neighborhood to probing the most distant galaxies in the early universe [2].

At the heart of this revolution are powerful observational facilities like NASA's Transiting Exoplanet Survey Satellite (TESS) [3], the Atacama Large Millimeter/submillimeter Array (ALMA) [4], ESO's High Acuity Wide-field K-band Imager (HAWK-I) [5], and the groundbreaking James Webb Space Telescope (JWST) [6]. Each of these instruments provides unique and complementary views of cosmic phenomena: TESS monitors stellar brightness variations to detect exoplanets [7], ALMA reveals the cold universe through submillimeter observations [8], HAWK-I delivers high-resolution near-infrared imaging [9], while JWST peers deeper into the infrared universe than ever before [10].

The true power of modern astronomy emerges when we combine data from these diverse instruments. A star-forming region observed by TESS in visible light, HAWK-I in near-infrared, and ALMA in millimeter wavelengths reveals different aspects of the star formation process that would remain hidden in any single observation [11]. Similarly, an exoplanet's atmosphere can be thoroughly characterized only by examining its transmission spectrum across multiple wavelength regimes [12].

However, this multi-instrument approach presents significant challenges. Astronomical datasets vary enormously in format, resolution, sensitivity, and spatial/spectral coverage [13]. Observations may be separated by years or even decades, requiring careful calibration and alignment [14]. The sheer volume of data - with modern surveys producing terabytes of information - demands efficient computational methods for processing and analysis [15].

This is where modern data science tools and techniques become essential. Python has emerged as the lingua franca of astronomical data analysis [16], with specialized libraries like Astropy for fundamental astronomy operations [17], Astroquery for accessing online databases [18], Pandas for data manipulation [19], and NumPy for numerical computations [20]. These tools enable researchers to overcome the technical barriers and extract meaningful scientific insights from complex, multi-wavelength datasets [21].

The ability to effectively combine and analyze observations from different instruments has opened new frontiers in nearly every area of astrophysics. Exoplanet researchers can now study planetary atmospheres in unprecedented detail [22]. Galactic astronomers can trace the lifecycle of stars from their birth in molecular clouds to their death in supernova explosions [23]. Cosmologists can observe galaxy formation and evolution across cosmic time [24]. In this new era of astronomy, the most exciting discoveries often come not from any single telescope, but from the synthesis of data across the electromagnetic spectrum [25].

With this said, we chose this topic because it’s rather interesting and has a lot of data. We chose these datasets due to their complementary strengths in various wavelengths and resolutions. For example, TESS offers high-cadence optical light curves that are essential for detecting exoplanets, whereas ALMA provides submillimeter observations of cold gas and dust. HAWK-I contributes high-resolution near-infrared imaging, and JWST enhances this with exceptional infrared sensitivity and spectroscopy. Collectively, these datasets enable us to investigate astronomical phenomena—such as star formation, exoplanet atmospheres, and galaxy evolution—from diverse viewpoints, thereby minimizing observational biases and facilitating a more thorough analysis.

2. Methodology

2.1. Objectives of This Study

The following study aims at cross-examining astronomical objects through the lenses of the different astronomical instruments presented.

Our primary objective relies on identifying objects target by all four facilities, within 1 degree of each other, based on their sky positions.

Afterwards, for objects observed by at least two facilities, we’ll be analyzing their sky coordinates to observe the discrepancies made by the instruments. Furthermore, we will complement this analysis by observing the temporal aspect of this data, alongside confirming the wavelength bias of each facility.

Key insights include detecting crowded fields or observational biases in sky positions, confirming temporal overlaps in potential coordinated studies and verifying complementary spectral range of each facility.

2.2. Experimental Setup

Given the complexity of this topic, we utilized AI support to determine the most appropriate analytical tools and to enhance the primary research questions that need to be explored.

Tools:
• Python 3.9+, Core programming language.
• Astroquery, API access to satellite archives.
• Astropy, Coordinate transformations, unit conversions.
• Pandas, Data wrangling and time-series analysis.
• NumPy/SciPy, Numerical computations and SED fitting.
• Networkx, Grahp Creation and manipulation.
• Matplotlib/Seaborn, Visualization and plotting.
• OS, Operation system functions.
• Defaultdic, Built-in containers.
• Rapidfuzz, Fast Spring Matching.
• Ydata-profiling, Provide an EDA in consistent and fast solution.
• Jiblib, Parallel processing.
• Sklearn, Provides of supervised and unsupervised learning algorithms.

Databases:

Astroquery Module	Archive	Data Products	URL
astroquery.mast	MAST (Mikulski Archive)	TESS, JWST	mast.stsci.edu
astroquery.alma	ALMA Science Archive	Interferometric visibilities, calibrated images	almascience.org
astroquery.eso	ESO Archive	HAWK-I	archive.eso.org

2.3. Data Cleaning

For all datasets, the cleaning plan was mostly conditioned by our objectives for this study. So, we decided to drop all columns that proved to be irrelevant to our study, only keeping the ones containing sky coordinates, times of data acquisition, spectral coverage and obviously the objects names. In addition to this, since the number of empty cells from each rows was neglible or even non-existant, we dropped those rows, and still kept a vast majority of the data.

Then we just performed more fine-grain tweaks to the data format in specific datasets.

- ALMA: em_min and em_max (millimeters) to em_min_nm and em_max_nm (nanometers).
- JWST, ALMA, TESS: All of this have times related to the beginning and end of data acquisition (t_min and t_max respectively) in MJD, which is a special calendar system that gives the number of days since November 17, of 1858. As for HAWKI, it contains the collum MJD-OBS, which can be interpreted as an instantaneous acquisition of data, also in MJD units. To normalize it we decided to assume an instantaneous capturing of data in the first three cited datasets, corresponding to the midpoint of the total time of data collection.

Finally, we only had one more cleaning step, which was deduplication of certain objects. In all datasets, we could have the same stellar object being identified in slightly different ways (e.g. Proxy Centauri vs PROXIMA CENTAURI). So, we had to use a few similarity metrics to trim the duplicated objects from each dataset.

2.4. Integration Plan

The objective of this plan is to facilitate the smooth deployment and interoperability of the Python code. Following the data cleaning process, all datasets will be standardized to include identical columns and metrics. Subsequently, these datasets will be merged, allowing for the creation of new columns derived from existing ones. For instance, the columns t_min and t_max will be combined to form a new column named MJD-OBS. Note that the schemas and integration model were made with already known columns to answer the question made. Also as said before on the last sentence of the data cleaning we approached the identity resolution to deduplicate certain objects using similarity metrics as jaccard distance, angular separation and absolute distance.

3. Results

To identify common targets across all four facilities, we created Figure 1. After analyzing the extensive results, we focused on the locations where all four facilities converged on the same target. This analysis revealed a total of ten distinct spots.

REFERENCES

[1] NASA Astrophysics Data System. (2023). "The Importance of Multi-Wavelength Astronomy." https://ui.adsabs.harvard.edu/

[2] European Space Agency. (2024). "Multi-Wavelength Observations in Modern Astrophysics." https://www.esa.int/Science_Exploration/Space_Science

[3] Ricker, G. R., et al. (2015). "Transiting Exoplanet Survey Satellite." Journal of Astronomical Telescopes, Instruments, and Systems.

[4] ALMA Observatory. (2022). "Science with ALMA." https://www.almaobservatory.org/en/science/

[5] ESO. (2021). "HAWK-I Instrument Overview." https://www.eso.org/sci/facilities/paranal/instruments/hawki/

[6] Gardner, J. P., et al. (2023). "The James Webb Space Telescope Mission." Nature Astronomy.

[7] NASA TESS Mission. (2023). "Exoplanet Discovery with TESS." https://tess.mit.edu/

[8] ALMA Science Portal. (2023). "Probing the Cold Universe." https://www.almaobservatory.org/en/science/

[9] ESO. (2020). "Near-Infrared Astronomy with HAWK-I." https://www.eso.org/public/teles-instr/paranal-observatory/vlt/vlt-instr/hawki/

[10] NASA JWST Science. (2023). "Infrared Astronomy with JWST." https://webb.nasa.gov/

[11] Luhman, K. L. (2022). "Star Formation Studies with Multi-Wavelength Data." Annual Review of Astronomy and Astrophysics.

[12] Madhusudhan, N. (2019). "Exoplanetary Atmospheres at Multiple Wavelengths." Space Science Reviews.

[13] National Research Council. (2021). "Challenges in Astronomical Data Analysis." https://nap.nationalacademies.org/

[14] Ivezic, Z., et al. (2019). "Statistics, Data Mining, and Machine Learning in Astronomy." Princeton University Press.

[15] LSST Science Collaboration. (2023). "Handling Large Astronomical Datasets." https://www.lsst.org/

[16] Greenfield, P. (2021). "Python in Astronomy." Astronomy & Computing.

[17] Astropy Collaboration. (2022). "The Astropy Project." Astronomy & Astrophysics.

[18] Ginsburg, A., et al. (2019). "Astroquery: Accessing Online Astronomical Data." Astronomical Journal.

[19] McKinney, W. (2017). "Pandas: Python Data Analysis Library." Journal of Open Source Software.

[20] Harris, C. R., et al. (2020). "Array Programming with NumPy." Nature.

[21] Price-Whelan, A. M., et al. (2022). "Astronomical Data Analysis in Python." Annual Review of Astronomy and Astrophysics.

[22] Seager, S. (2023). "Exoplanet Atmosphere Characterization." Reports on Progress in Physics.

[23] McKee, C. F., & Ostriker, E. C. (2022). "Theory of Star Formation." Annual Review of Astronomy and Astrophysics.

[24] Madau, P., & Dickinson, M. (2021). "Cosmic Star Formation History." Annual Review of Astronomy and Astrophysics.

[25] NASA. (2023). "The Future of Multi-Messenger Astronomy." https://science.nasa.gov/astrophysics/multi-messenger-astronomy

Attachments

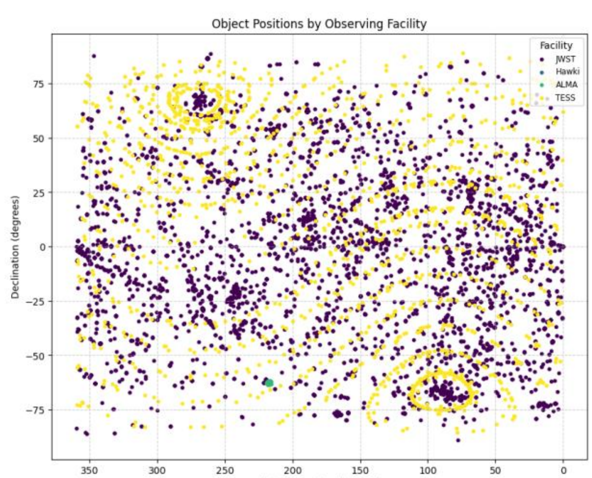

Figure 1: Object Positions by Observing Facility.