Practice 3

Astroinformatics I

Semester 1, 2025

Ana Pereira Aguirre

June 30, 2025

1 Introduction

This report contains the solutions of the Practical III of the Astroinformatics I course. The tasks consist of manipulating light curve files using `python` scripts. All the code and explanations are developed below.

2 Exercise solutions

2.1 Task 1

Repeat the plots from graded practice 1, but now with Python using the light curve files from practice 2. For the plots, take into account how to make them more readable.
To plot the light curves with `python`, we first import the following libraries

1. import pandas as pd: For reading and manipulating CSV data tables.
2. import matplotlib.pyplot as plt: For plotting the light curves.
3. from astropy.stats import sigma_clip
4. import glob: For finding all files that match a filename pattern
5. import re: For extracting numbers from filenames

Then, we use a for loop to iterate over these files and plot each light curve individually in a single execution of the code:

```python
import pandas as pd
import matplotlib.pyplot as plt
from astropy.stats import sigma_clip
import glob
import re

tess_lc = glob.glob("tess_lc_*.csv")


for filename in tess_lc:
    # to read the csv file
    df = pd.read_csv(filename)

    # remove nan values
    df = df[["TIME", "PDCSAP_FLUX", "PDCSAP_FLUX_ERR"]].dropna()

```

```python
    # Identify outliers with sigma clipping
    s_clip = sigma_clip(df["PDCSAP_FLUX"], sigma_upper=3, sigma_lower=None)
    mask = s_clip.mask
    normal = df[~mask]
    outliers = df[mask]

    # Graph
    plt.figure(figsize=(10, 5))
    plt.errorbar(normal["TIME"], normal["PDCSAP_FLUX"], yerr=normal["PDCSAP_FLUX_ERR"],
                 fmt='.', alpha=0.4, color="dimgrey")

    plt.errorbar(outliers["TIME"], outliers["PDCSAP_FLUX"], yerr=outliers["PDCSAP_FLUX_ERR"],
                 fmt='.', alpha=0.7, color="palevioletred", label="Outliers")

    # Extract the number from the filename to put in plot the title
    match = re.search(r'tess_lc_(\d+)', filename)
    name = match.group(1) if match else filename

    plt.title(f"TESS Light Curve {name}")
    plt.xlabel("Time - 2457000 [BJD days]")
    plt.ylabel("Flux (PDCSAP) [e-/s]")
    plt.legend()
    plt.grid(False)
    plt.tight_layout()
    plt.savefig(filename.replace(".csv", ".png"))
    plt.show()
```

2.2 Task 2

When you make the plots, can you identify outliers? Highlight them. Try writing code to identify at least the most extreme outliers.
To identify the outliers, we use the `sigma_clip` function from the `astropy` library. A data point is flagged as an outlier if it satisfies the condition:

$$|x - \mu| > N \cdot \sigma$$

Where:

- μ is the mean
- σ is the standard deviation
- N is the sigma threshold (in this case, $N = 3$).

So, to find the outliers we write the following line of code:

```
# Identify outliers with sigma clipping
s_clip = sigma_clip(df["PDCSAP_FLUX"], sigma_upper=3, sigma_lower=None)
mask = s_clip.mask
normal = df[~mask]
outliers = df[mask]
```

When we run the code, we obtain:

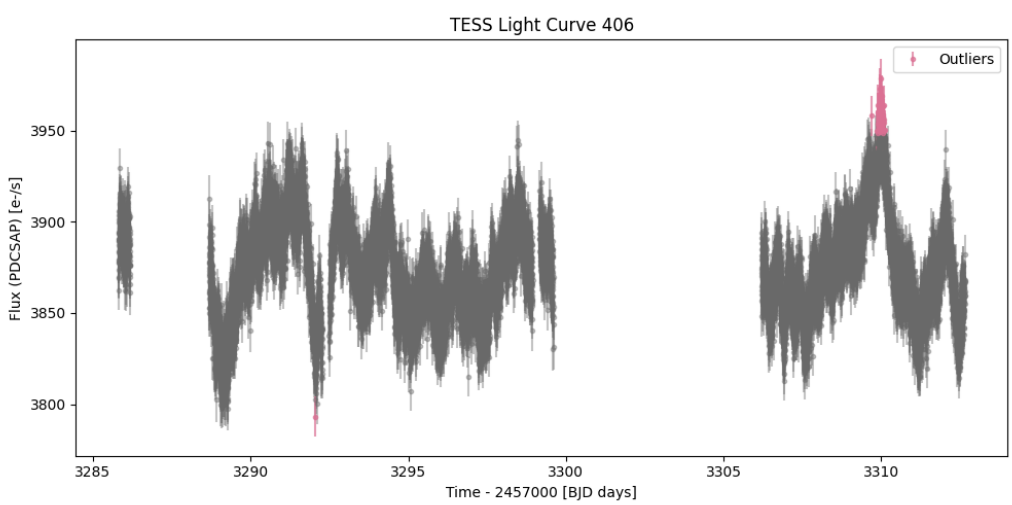


Figure 1: TESS Light Curve 1755406

2.3 Task 3

Think about basic statistics to describe light curves, such as amplitudes, and implement at least two of them.
To describe the light curves, we computed several basic statistical measures. These allow us to characterize the general behavior and variability of each light curve:

1. Amplitude: the difference between the maximum and minimun flux values.
2. Mean: the average flux value.
3. Standar deviation: a measure of how much the flux varies around the mean.
4. Median
5. Skewness: indicates whether the distribution of flow values is symmetrical or skewed towards higher or lower values.
6. Kurtosis: describes whether the flux distribution has pronounced peaks or is flat, as compared to a normal distribution.

To obtain these elements, the following code was developed:

```
# Import libraries
import pandas as pd

lightcurves = ["tess_lc_696.csv", "tess_lc_736.csv"]

# to save the results
results = []

for lc in lightcurves:
    df = pd.read_csv(lc)

    # remove nan values
    flux = df["PDCSAP_FLUX"].dropna()

    # Stats calculation
    stats = {
        "name": lc,
        "amplitude": flux.max() - flux.min(),
        "mean": flux.mean(),
        "median": flux.median(),
```

```
        "standard deviation": flux.std(),
        "skewness": flux.skew(),
        "kurtosis": flux.kurtosis()
    }

    results.append(stats)

# to save as csv file
df_stats = pd.DataFrame(results)
df_stats.to_csv("stats_lc.csv", index=False)

df_stats
```

Where we analyze two light curves and applying the code we obtain:

Name	Amplitude	Mean	Median	Std. Dev.	Skewness	Kurtosis	Estimated Period
tess_lc_696.csv	953.507	19252.623	19250.094	214.281	0.0943	-1.0859	3.3148
tess_lc_736.csv	137.562	15060.933	15061.028	16.308	-0.0027	0.0347	0.0014

Table 1: Basic statistics for the light curves.

TESS Light Curve 02104696

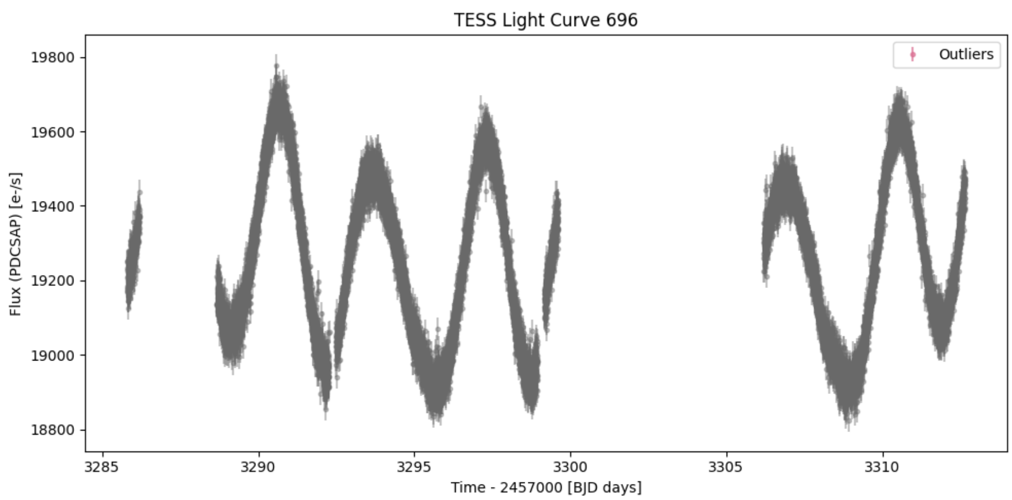


Figure 2: TESS Light Curve 02104696

The light curve corresponding to the `tess_lc_696.csv` file has a high amplitude of 953.507, which is evidence of significant variability in source brightness. The mean and median of the data (19252.623 and 19250.094, respectively) are similar, indicating a relatively symmetric distribution. The estimated period of variation is 3.3148 days, which points to a periodic modulation. According to the general shape of the curve, it is possible that it is a variable star, since the characteristic morphology of a transiting exoplanet, which usually shows short and periodic drops in the flux, is not appreciated.

TESS Light Curve 01950736

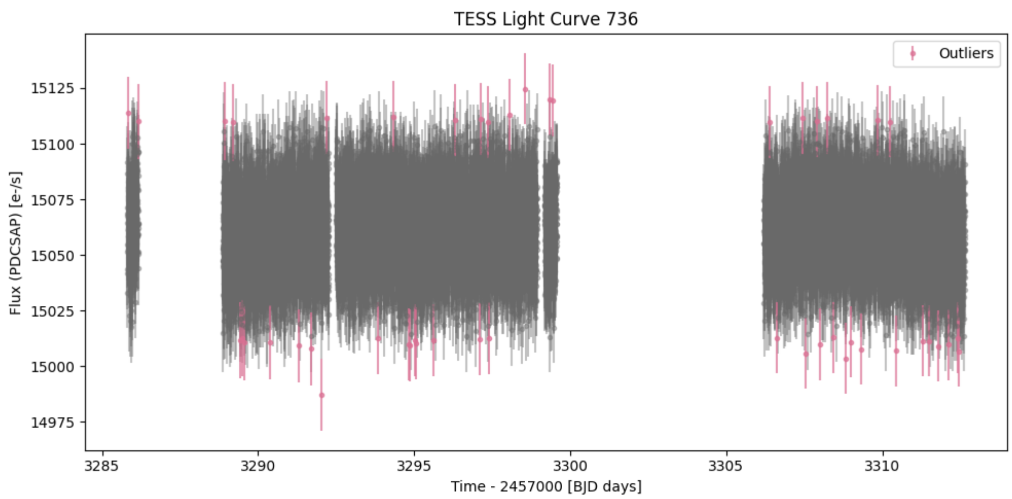
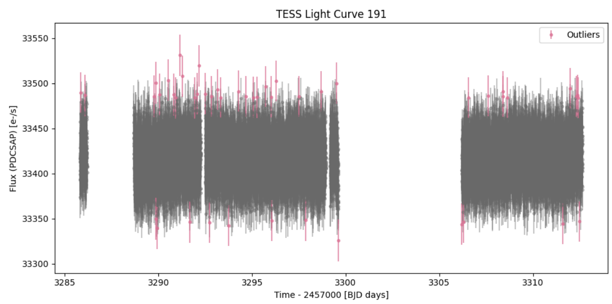


Figure 3: TESS Light Curve 01950736

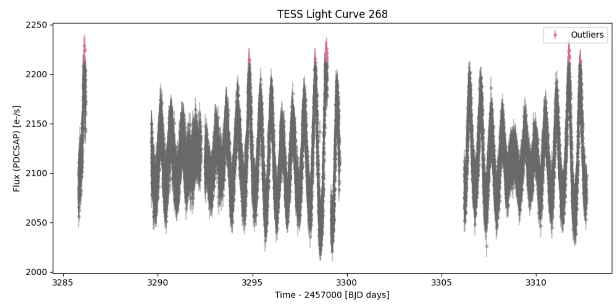
The light curve corresponding to the `tess_lc_736.csv` file shows a low amplitude of 137.562, so we can say that the observed source shows little variability in its brightness over time. The mean (15060.933) and median (15061.028) are nearly equal, indicating a very symmetric distribution. The estimated period is only 0.0014 days. This type of light curve may be characteristic of exoplanet transits, since it presents small dips in brightness that could correspond to the passage of a planet in front of its star.

A Additional Light Curves

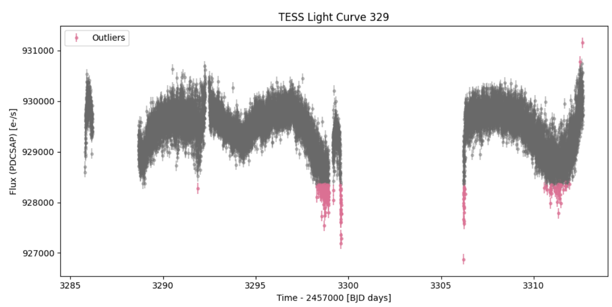
The following figures present the additional light curves generated from TESS data in Sector 73, as part of the data processing and visualization activity. Each plot display the flux variation of a selected object over time, using `PDCSAP_FLUX` values.



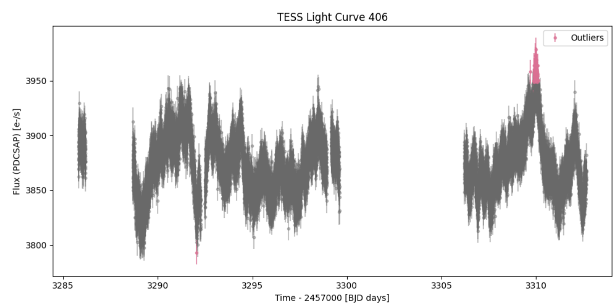
((a)) Light curve of TESS object 191.



((b)) Light curve of TESS object 268.

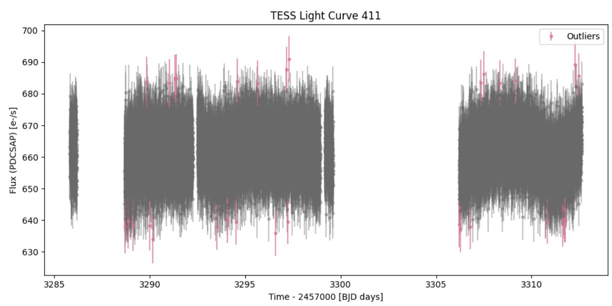


((c)) Light curve of TESS object 329.

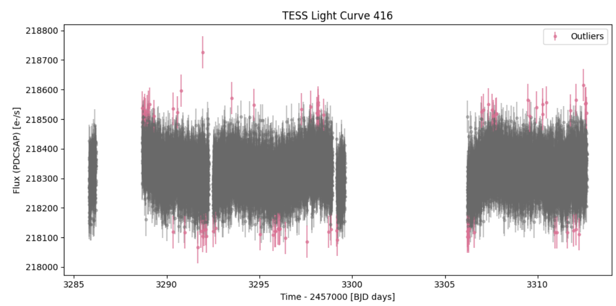


((d)) Light curve of TESS object 406.

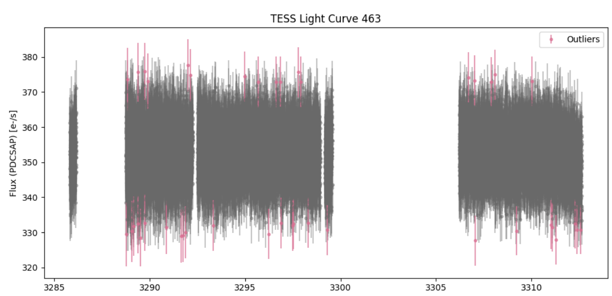
Figure 4: Light curves of selected TESS objects. Each plot shows the flux variability over time, with the x-axis representating `BTJD` and y-axis respresentating the normalized `PDCSAP_FLUX`, with errors.



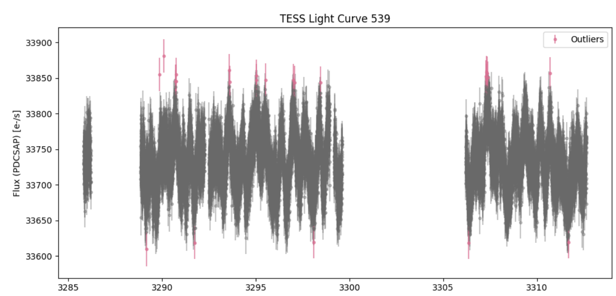
((a)) Light curve of TESS object 411.



((b)) Light curve of TESS object 416.

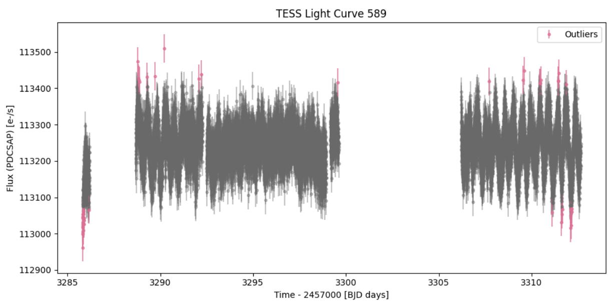


((c)) Light curve of TESS object 463.

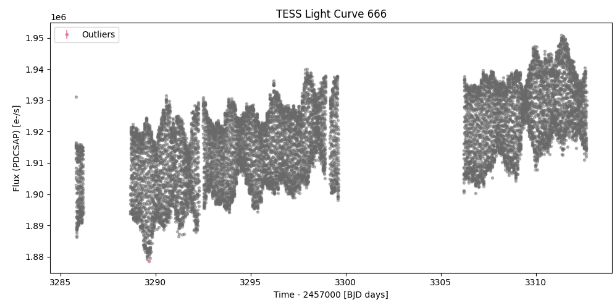


((d)) Light curve of TESS object 539.

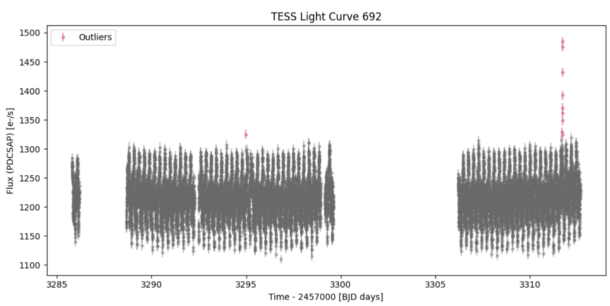
Figure 5: Light curves of selected TESS objects. Each plot shows the flux variability over time, with the x-axis representating `BTJD` and y-axis respresentating the normalized `PDCSAP_FLUX`, with errors.



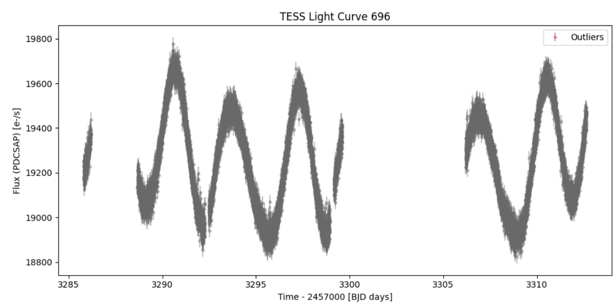
((a)) Light curve of TESS object 589.



((b)) Light curve of TESS object 666.

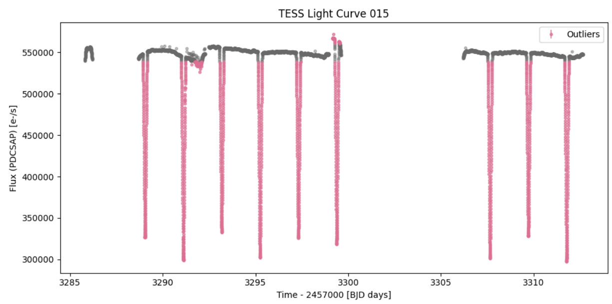


((c)) Light curve of TESS object 692.

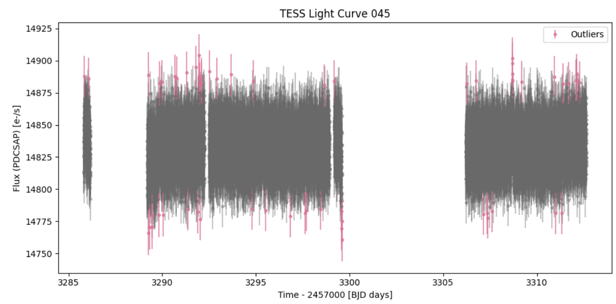


((d)) Light curve of TESS object 696.

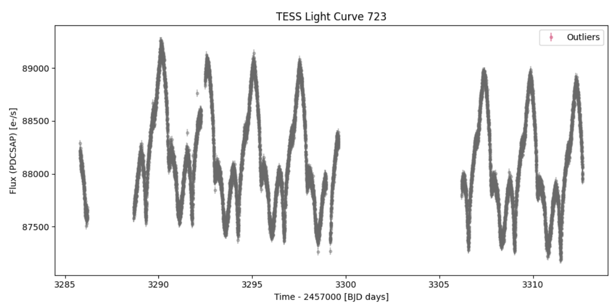
Figure 6: Light curves of selected TESS objects. Each plot shows the flux variability over time, with the x-axis representating `BTJD` and y-axis respresentating the normalized `PDCSAP_FLUX`, with errors.



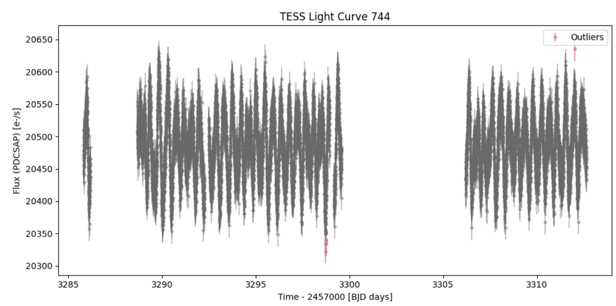
((a)) Light curve of TESS object 015.



((b)) Light curve of TESS object 045.

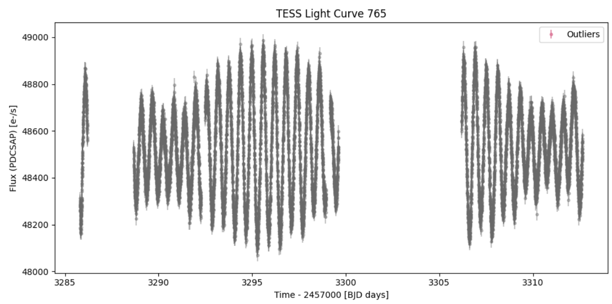


((c)) Light curve of TESS object 723.

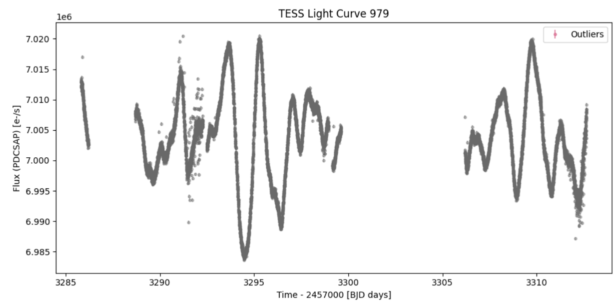


((d)) Light curve of TESS object 744.

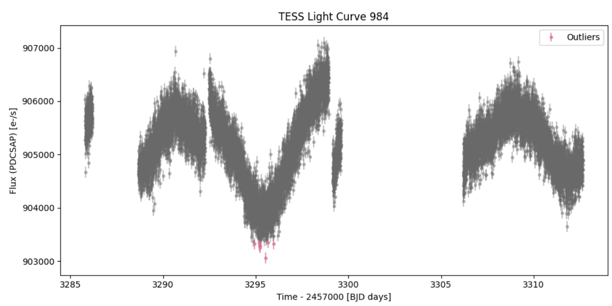
Figure 7: Light curves of selected TESS objects. Each plot shows the flux variability over time, with the x-axis representating `BTJD` and y-axis respresentating the normalized `PDCSAP_FLUX`, with errors.



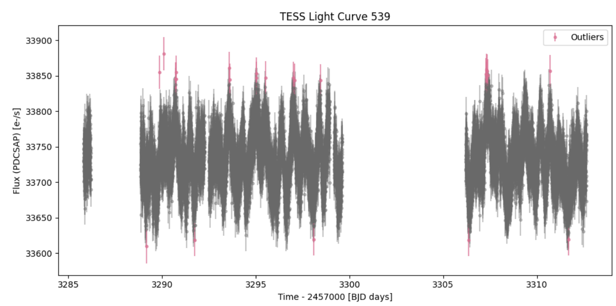
((a)) Light curve of TESS object 765.



((b)) Light curve of TESS object 979.



((c)) Light curve of TESS object 984.



((d)) Light curve of TESS object 539.

Figure 8: Light curves of selected TESS objects. Each plot shows the flux variability over time, with the x-axis representating `BTJD` and y-axis respresentating the normalized `PDCSAP_FLUX`, with errors.