EDA*

Yufei Liu

2025-02-09

1   Data summary

In this project, we are interested in detecting stellar flares using different methods and compare the model performance for three stars using TESS light curve data for TIC 0131799991, TIC 129646813, and TIC 031381302. We mainly focused on two variables from the dataset, time and PDCSAF Flux, which refers to the corrected flux after summing the calibrated pixels within the TESS optimal photometric aperture. The original dataset, although, includes missing values for both flux and time, which could be a problem when we modeling the data.

For the following exploratory data analysis, I will use data for TIC 013799991 as an example, and will expand the implements to other two stars.

It is noticeable for TIC 013799991 that it actually has a missing period between time index 1529 to 1535 as shown in Table 1, which may be the result of instrumental errors. For the missing values in flux, as shown in Figure 1b, the heatmap shows that only the very first a few observations are missing (observation in light blue). In this case, since the time in the data refers to the Barycentric Julian Date, instead of converting it to a `ymd` variable, I kept the original numeric number since I found it is unnecessary to convert and may introduce extra errors in the process. So the heatmap did not show the missing period in time.
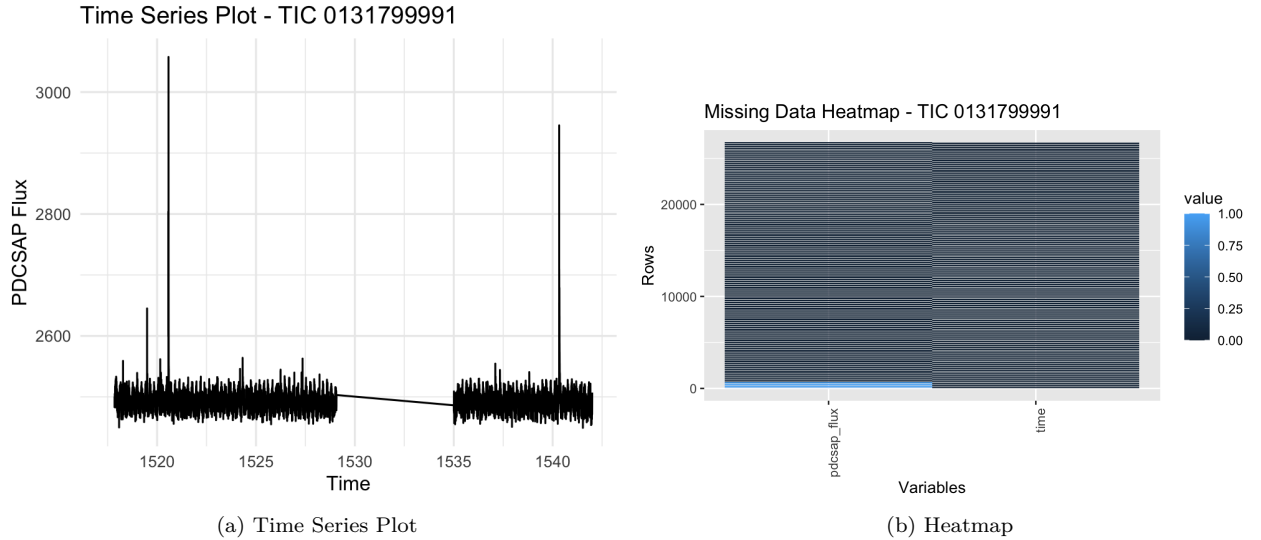

(a) Time Series Plot

(b) Heatmap

Figure 1: Visualization of time series TIC 0131799991

---

*Code and data are available at: https://github.com/Florence-Liu/Stellar-Flare

Table 1: Part of data with missing period

Index	Time	PDCSAF Flux
8343	1529.065	2484.593
8344	1529.067	2478.021
8345	1529.068	2503.103
8346	1535.001	2486.231
8347	1535.003	2476.867
8348	1535.004	2486.383
8349	1535.006	2487.144
8350	1535.007	2479.111
8351	1535.008	2488.168

Since it is a time series data, I conducted the time series analysis to check some features of the dataset, including the ACF, PACF, and seasonal trend decomposition.
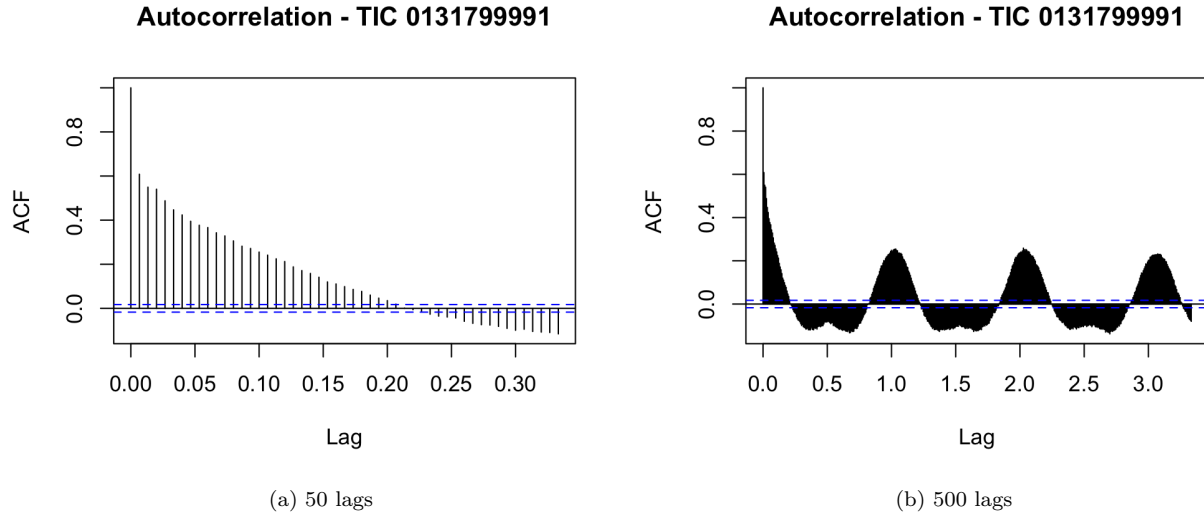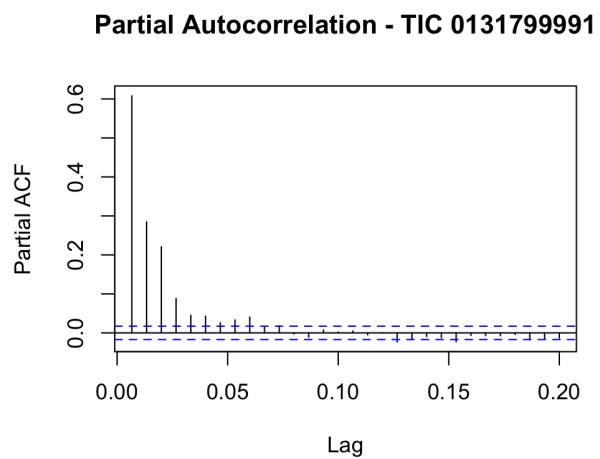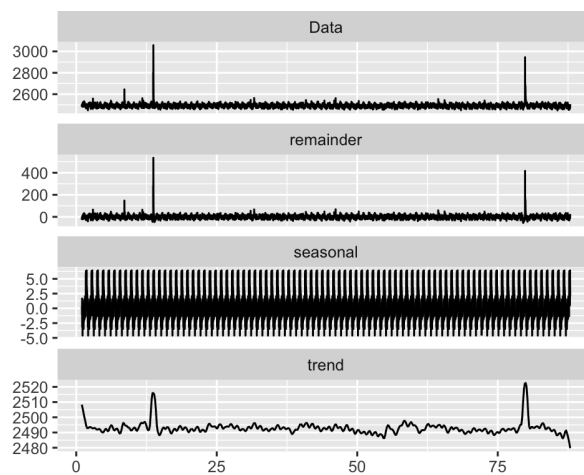

(a) 50 lags

(b) 500 lags

Figure 2: ACF for time series TIC 0131799991 with different maximum lag

(a) PACF

(b) STL decompositioin

Figure 3: Result for PACF and STL decomposition

2   Handling missing data