Measuring Significant Increase or Decrease in The Flux of Star KIC 8462852 in The Constellation of Cygnus

Abdulhalim Hamada, Ekin Oran, Ryan Lam

University of Waterloo

Introduction

KIC 8462852 is a yellow-white dwarf star that is located in the constellation of Cygnus, being 1470 light-years away from our solar system. It was discovered in 2015 by Tabetha Boyajian and was named after him as Tabby's star. Its brightness happens to be changing irregularly, having irregular periods of dimming. It does not follow a particular time or the depth of the dipping changes. The maximum dips have been as deep as 20 percent. At first, scientists thought it might be an exoplanet but the dimming of the star is so great that it rules that option out. Usually, when a planet passes between two heavenly bodies, it dims by 1 percent or less at regular intervals. Also, it is observed that when the star dims, some wavelengths seem to be more blocked than others. Hence it cannot be another solid object causing it to dim because solid objects would block all wavelengths equally. A team of astronomers has been studying the star since 2016 and came up with the conclusion that there might be a cloud of thin dust and debris caused by a binary companion star. In 2020, the astrometric data from the Gaia satellite agreed with the team's finding and included a faint star that's very close to KIC 8462852, separated by 880 au. The binary companion is a red dwarf with the name KIC 8462852 B. The astronomers believe this companion star might cause fluctuations in the star's brightness.

Data were collected in four orientations as the telescope rolled every 93 days. After a month of collecting data continuously, FFI images are the last step of the observation before it downlinks to Earth. A single FFI image contains the full set of all science and collateral pixels using the CCDs of a camera. The images were obtained every 30 minutes during the Kepler mission. It takes approximately 1 day to resume the data collection after taking FFI images. We then use these images to collect all the flux from the star. 11x11 pixel apertures are created in the center of the brightest pixel of the star. Any star with a standard deviation of the residuals that are larger than 0.5% is discarded. Then, the flux level of KIC 8462852 is determined by comparing its aperture photometry to the aperture photometry recorded from the calibrator stars.

Data, Calculations and Analysis

Tables and Calculations

Graph 1

For categorical analysis associated with the frequency of the categorical variable, we use the one-way chi squared statistical test. Before even stating our null hypothesis we first construct a frequency table for the orientation (categorical) variable.

Orientation	Frequency
0	11
1	12
2	12
3	52
Grand Total	**87**

Our null hypothesis here would be that the orientation (of how the telescopes were kept) frequencies were random and if our p value is less than 0.05 (5% confidence) then we could reject this notion and assume that there is a bias when it comes to the distribution of KIC 8462852 flux and its orientation.

Here is what we get we apply scipy.stats to our data:

```
In [1]: #1
import scipy.stats as stats
orientation_freq=[11,12,12,52]
chi2_statistic, chi2_pVal=stats.chisquare(orientation_freq)

if chi2_pVal<0.05:
    print("The null hypothesis IS  rejected because",chi2_pVal," is smaller than 0.05 concluding that there is a bias and the",
"orientation of the telescope is not random")
else:
    print("The null hypothesis IS NOT  rejected because",chi2_pVal,"is greater than 0.05 concluding that the orientation of",
"the telescope is random")

The null hypothesis IS  rejected because 3.9478783826185384e-12  is smaller than 0.05 concluding that there is a bias and the o
rientation of the telescope is not random
```

Rejecting the null hypothesis when just looking at the distribution of the orientations frequency we can tell that the orientations are not just random, there is bias involved in the distribution of the orientations of the telescope and this can be looked at as a source of error.

Now we want to conduct a one sample t-test (using the first value of uncertainty to be the population mean as we want to see if it will deviate from this) on the uncertainty to see if the uncertainty is close to being constant for every flux and orientation or does it have a bias on the flux (the uncertainty changes for different flux).

Here our null hypothesis would be that there is no change in the uncertainty (does not change significantly) as the flux changes.

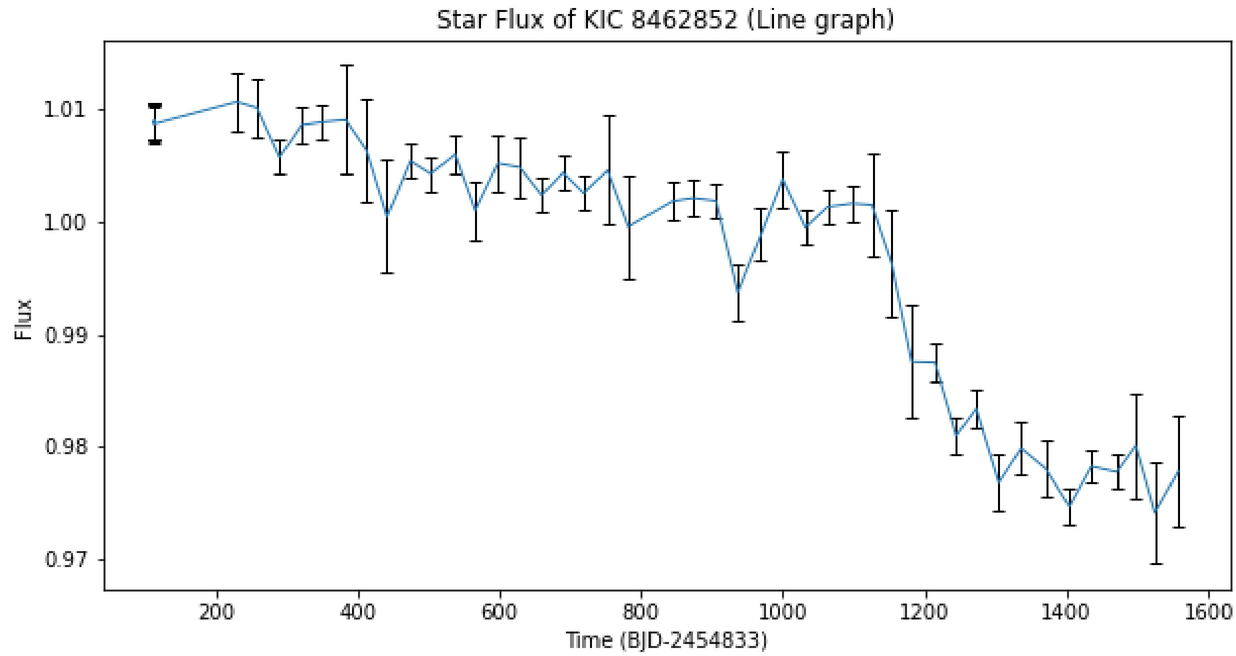
```
In [4]: import numpy as np
data = np.loadtxt("finaldataset.csv", delimiter=",", skiprows=1)
time = data[:,0]
flux = data[:,1]
uncert = data[:,2]
trueUncert=data[0,2]

stat_uncert, pval_uncert = stats.ttest_1samp(uncert, trueUncert)

if pval_uncert<0.05:
    print("The null hypothesis IS  rejected because",pval_uncert," is smaller than 0.05 concluding that there is a bias between",
"the uncertainty and the flux and the uncertainty of the flux is not close to being constant")
else:
    print("The null hypothesis IS NOT  rejected because",pval_uncert,"is greater than 0.05 concluding that the uncertainty of",
"the flux is close to being constant")
```

```
The null hypothesis IS  rejected because 8.031277241130186e-06  is smaller than 0.05 concluding that there is a bias between th
e uncertainty and the flux and the uncertainty of the flux is not close to being constant
```

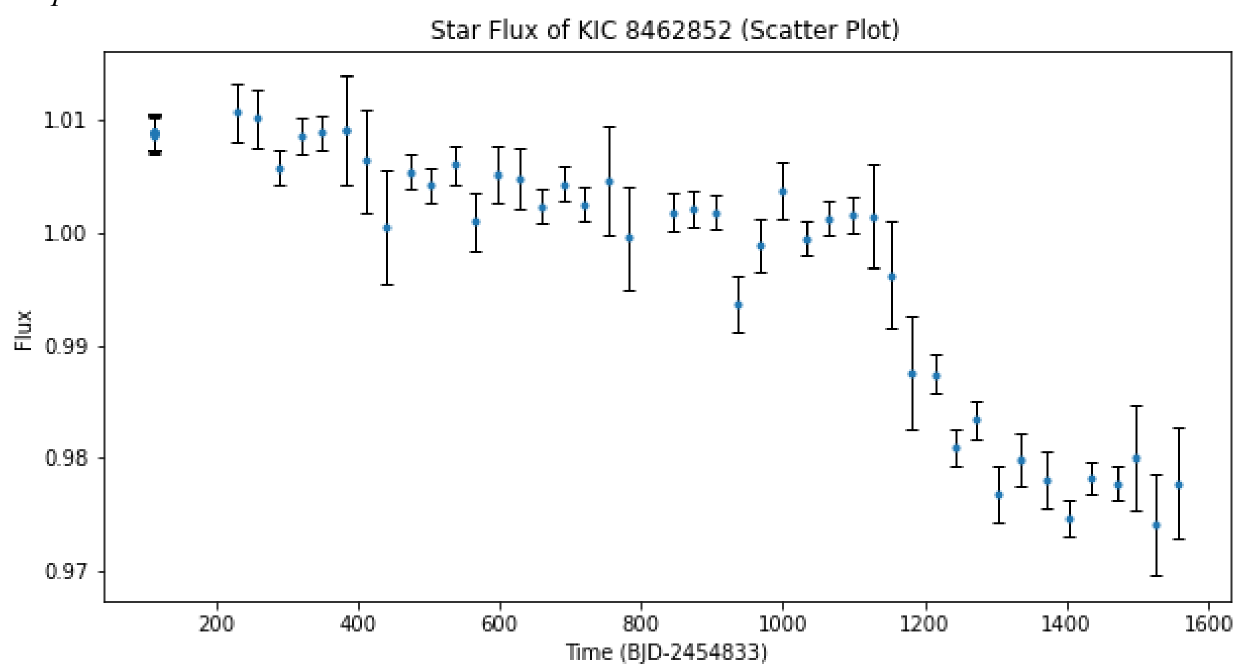
The null hypothesis is rejected and our alternative hypothesis, the uncertainty changes significantly for a change in flux, is accepted.

Graphs and Analysis

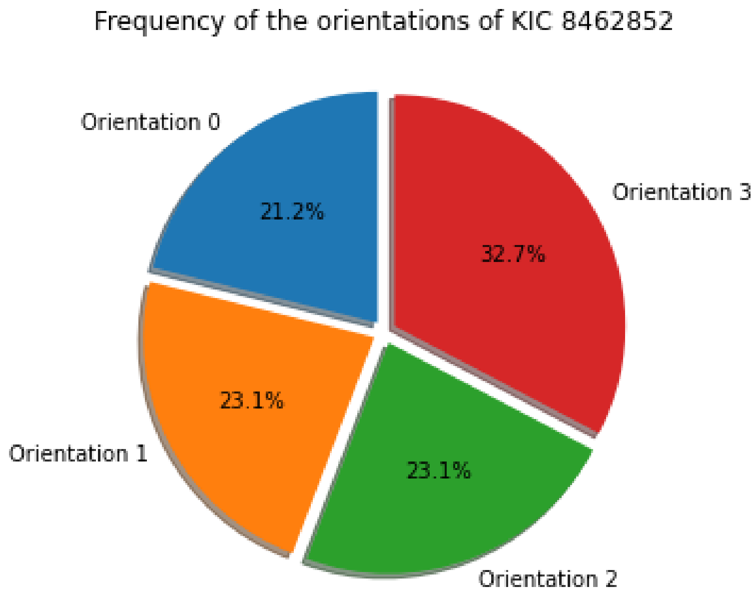
Graph 1



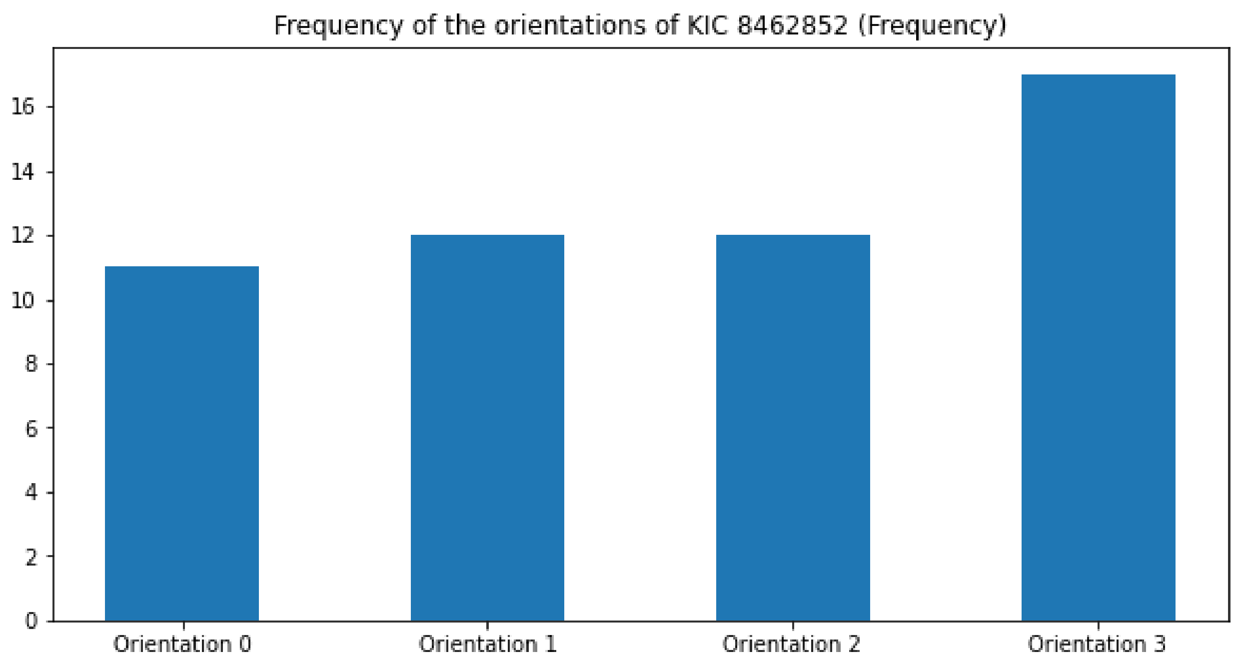
Graph 2



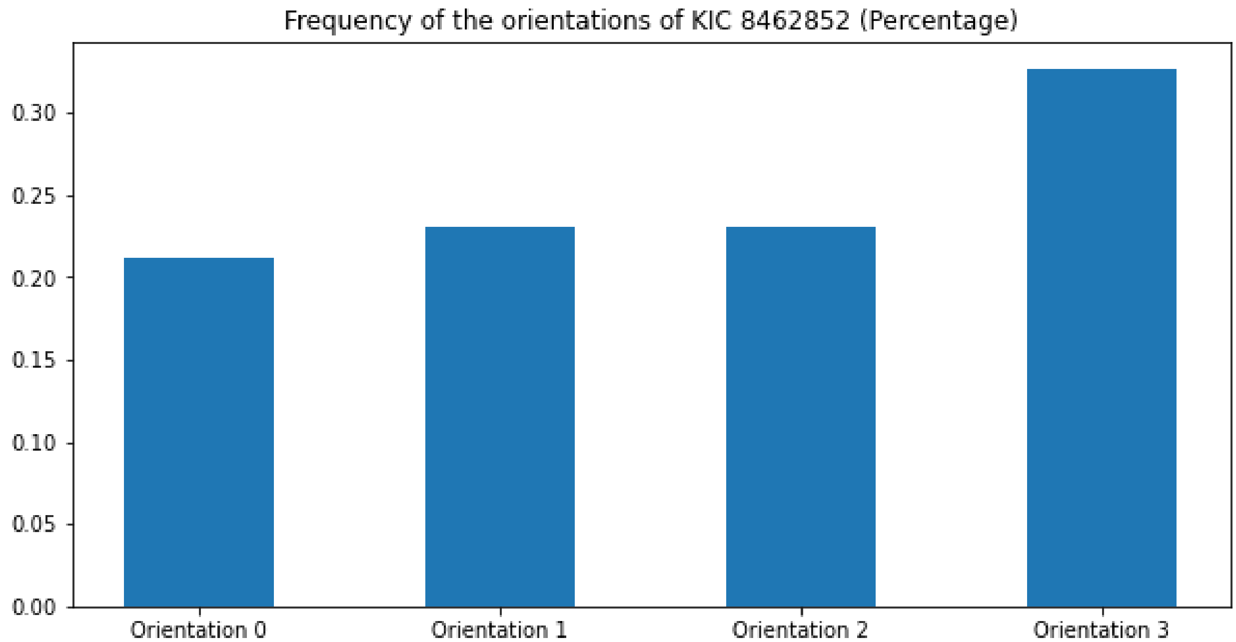
Graph 3



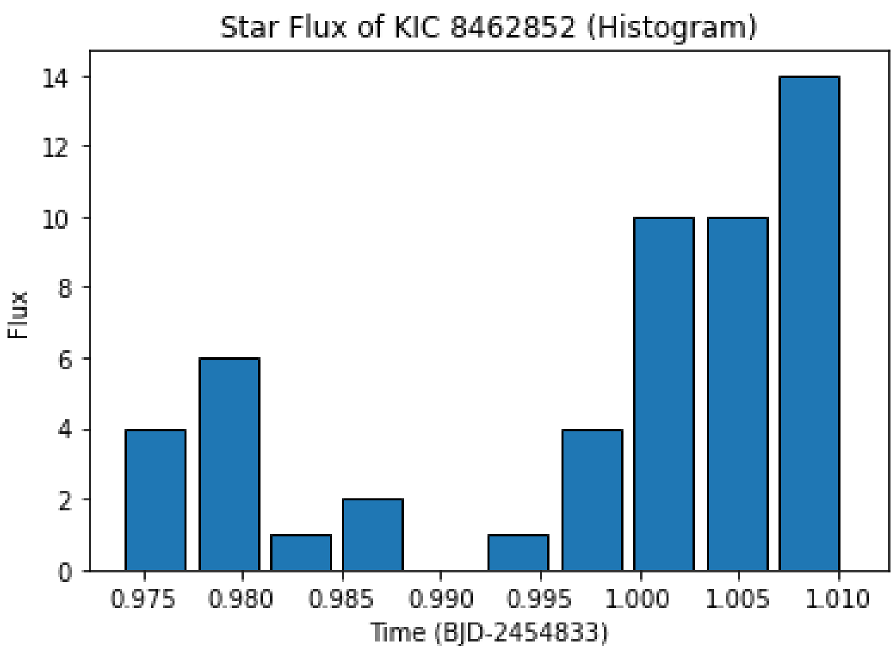
Graph 4



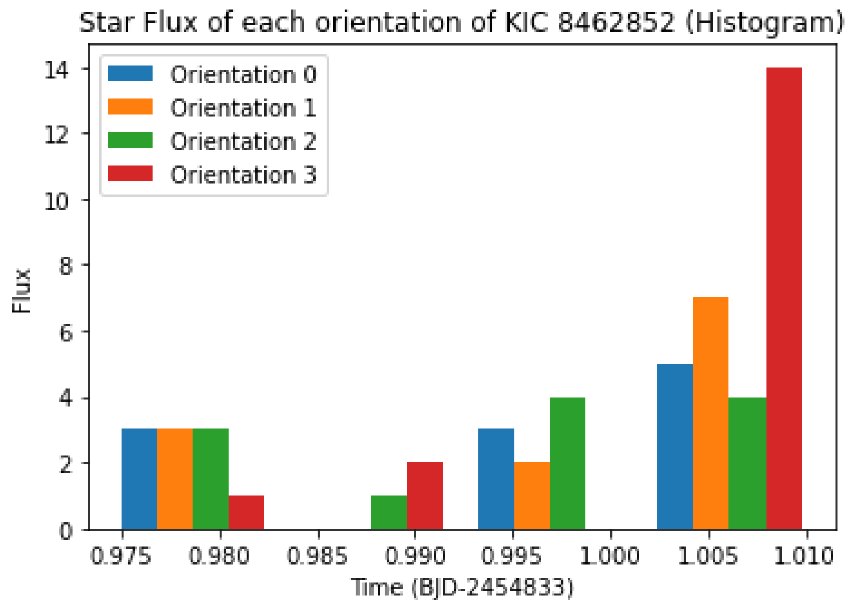
Graph 5



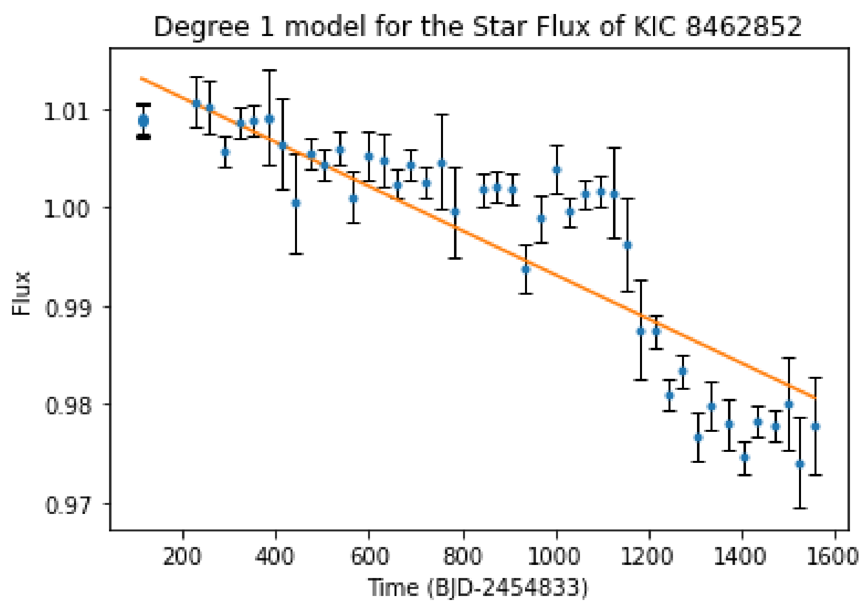
Graph 6



Graph 7

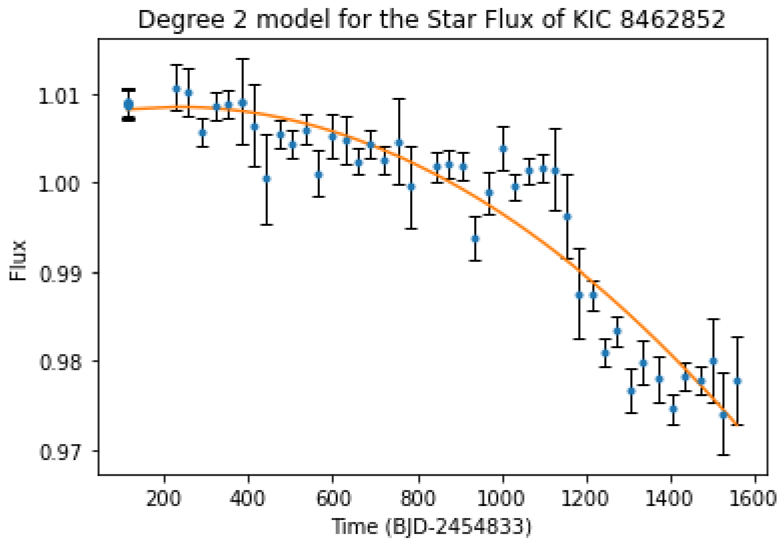


Graph 8



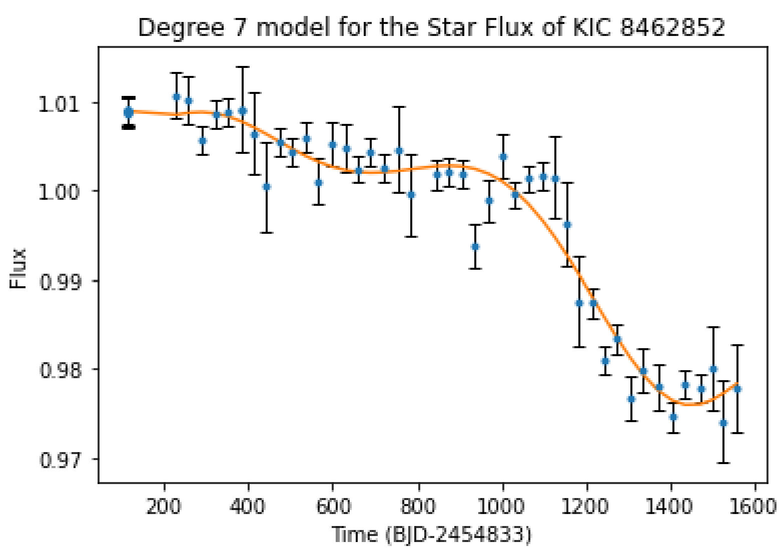

```
For degree 1, R^2 is: 0.7993336922893973
```

Graph 9



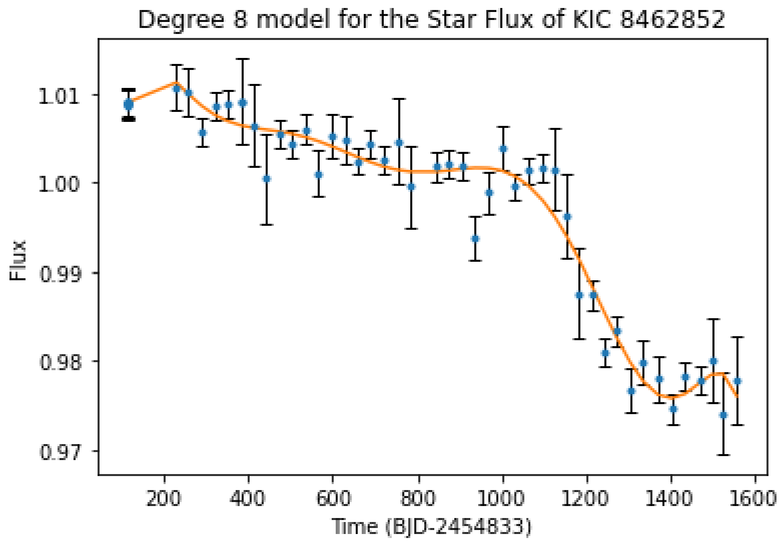
For degree 2, R^2 is: 0.8980902144397775

Graph 10



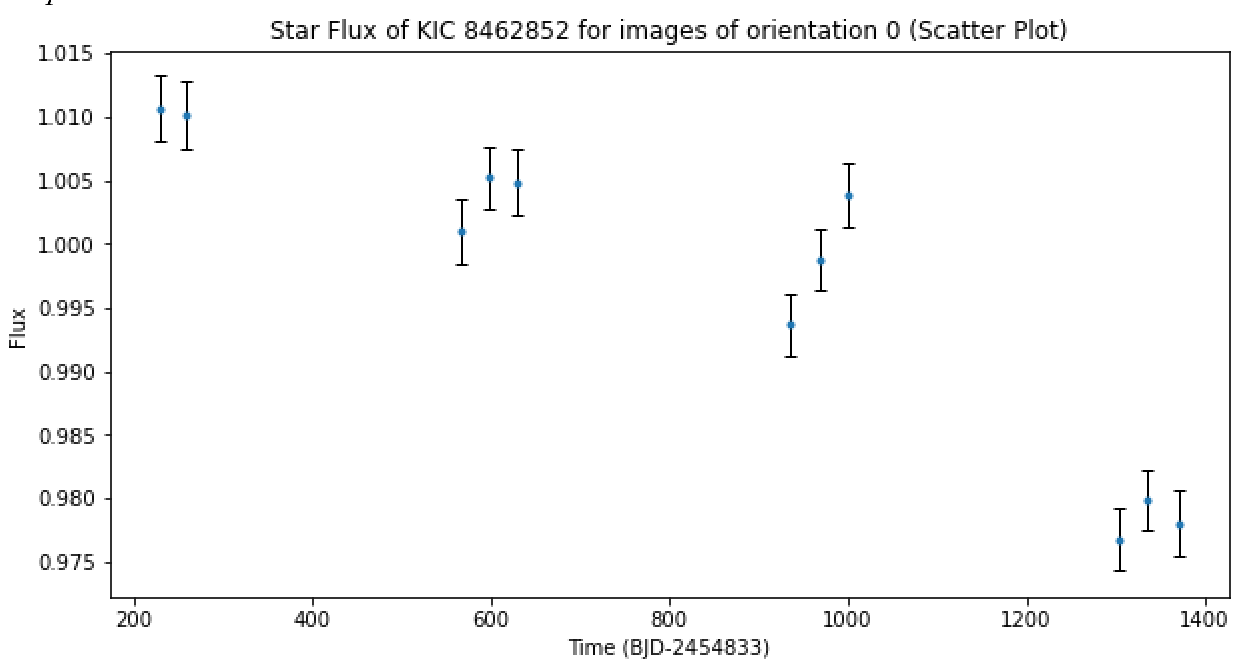
For degree 7, R^2 is: 0.9463885025664199

Graph 11



For degree 8, R^2 is: 0.955507457020038

Graph 12

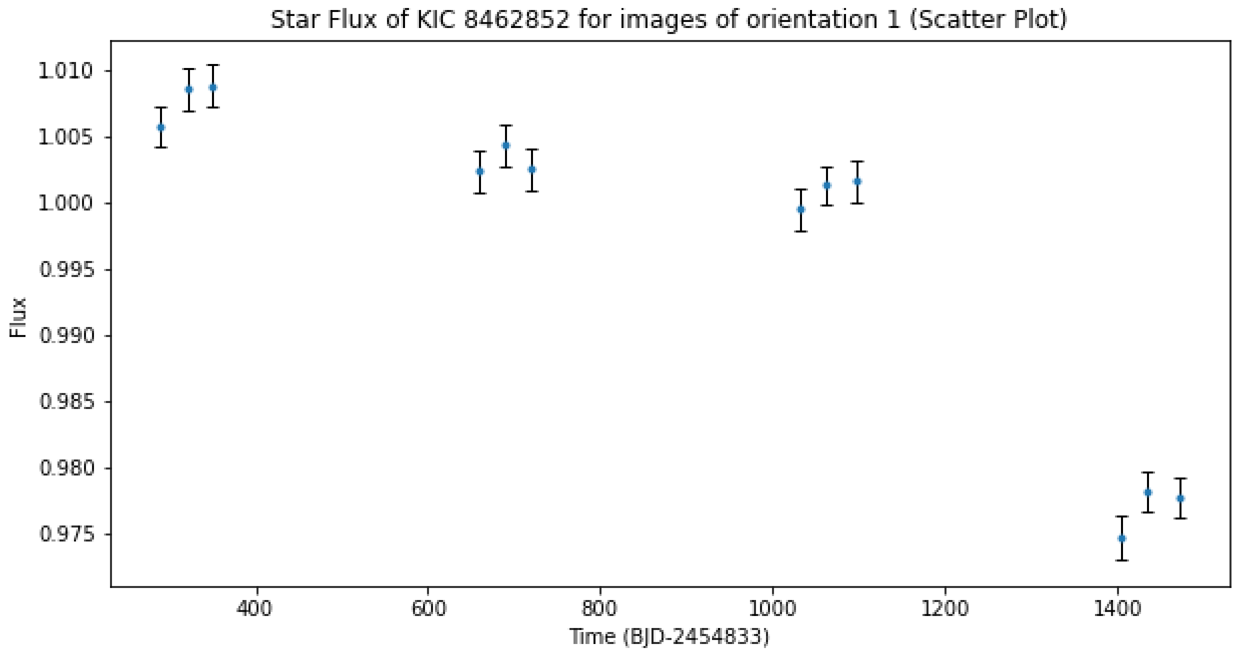


We confirm the KIC 8462852 data/images of orientation 0 IS distributed normally with a p-value of 0.34087553456930275 and a statistic of 2.152475739181673

The null hypothesis is that the flux of KIC 8462852 for images of orientation 0 is not decreasing. The alternative hypothesis is that the flux of KIC 8462852 for images of orientation 0 is decreasing. For this hypothesis test, we will use a 1 sample t-test because it compares the mean of the datapoints to a theoretical mean. The theoretical mean in this test would be the first value/datapoint of the orientation 0 images. This is because if flux is decreasing, then the mean of the orientation 0 images data will be significantly away from the first value/datapoint of the orientation 0 images

We confirm that the flux of KIC 8462852 for images of orientation 0 IS decreasing with a p-value of 0.004442449769574793 and a statistic of -3.6526943293840453

Graph 13

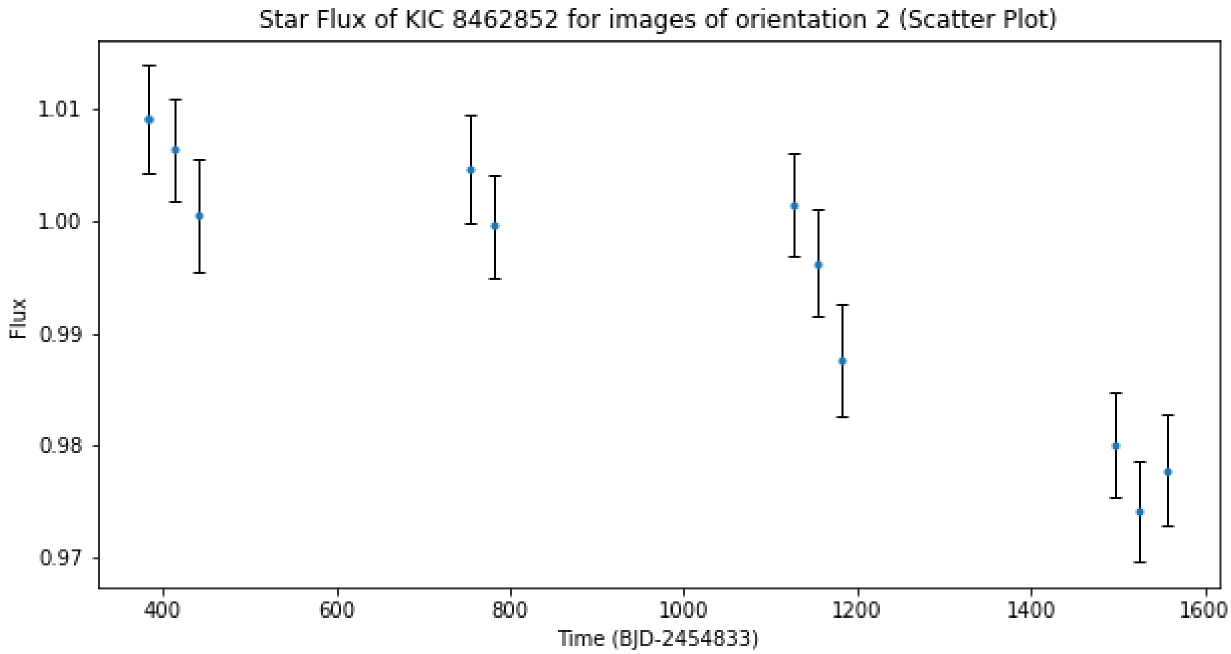


We confirm the KIC 8462852 data/images of orientation 1 IS distributed normally with a p-value of 0.1977024491962826 and a statistics of 3.2419843208674606

The null hypothesis is that the flux of KIC 8462852 for images of orientation 1 is not decreasing. The alternative hypothesis is that the flux of KIC 8462852 for images of orientation 1 is decreasing. For this hypothesis test, we will use a 1 sample t-test because it compares the mean of the datapoints to a theoretical mean. The theoretical mean in this test would be the first value/datapoint of the orientation 1 images. This is because if flux is decreasing, then the mean of the orientation 1 images data will be significantly away from the first value/datapoint of the orientation 1 images

We confirm that the flux of KIC 8462852 for images of orientation 1 IS decreasing with a p-value of 0.03686293925839836 and a statistic of -2.3744187203365827

Graph 14

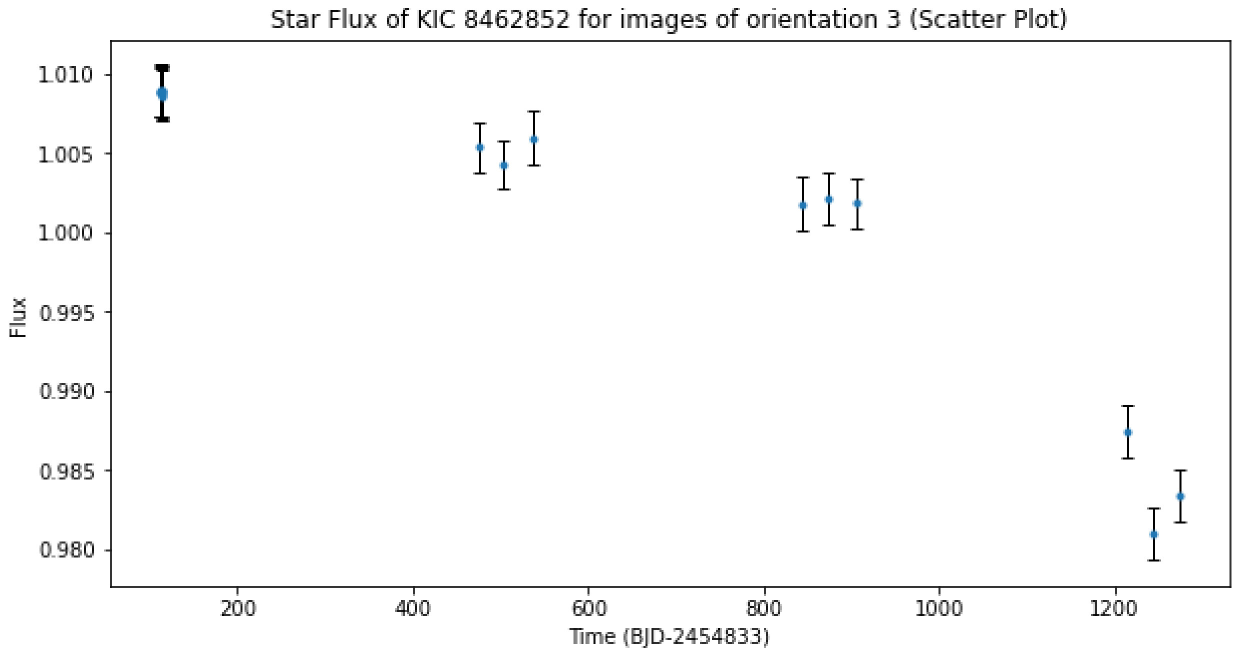


We confirm the KIC 8462852 data/images of orientation 2 IS distributed normally with a p-value of 0.34870033087202074 and a statistics of 2.1070847528559318

The null hypothesis is that the flux of KIC 8462852 for images of orientation 2 is not decreasing. The alternative hypothesis is that the flux of KIC 8462852 for images of orientation 2 is decreasing. For this hypothesis test, we will use a 1 sample t-test because it compares the mean of the datapoints to a theoretical mean. The theoretical mean in this test would be the first value/datapoint of the orientation 2 images. This is because if flux is decreasing, then the mean of the orientation 2 images data will be significantly away from the first value/datapoint of the orientation 2 images

We confirm that the flux of KIC 8462852 for images of orientation 2 IS decreasing with a p-value of 0.003189276596626131 and a statistics of -3.7537151317658846

Graph 15



We confirm the KIC 8462852 data/images of orientation 3 IS distributed normally with a p-value of 0.348700330872020074 and a statistics of 2.1070847528559318

The null hypothesis is that the flux of KIC 8462852 for images of orientation 3 is not decreasing. The alternative hypothesis is that the flux of KIC 8462852 for images of orientation 3 is decreasing. For this hypothesis test, we will use a 1 sample t-test because it compares the mean of the datapoints to a theoretical mean. The theoretical mean in this test would be the first value/datapoint of the orientation 3 images. This is because if flux is decreasing, then the mean of the orientation 3 images data will be significantly away from the first value/datapoint of the orientation 3 images.

We confirm that the flux of KIC 8462852 for images of orientation 3 IS decreasing with a p-value of 0.012886407828021116 and a statistics of -2.7983025417722343

Discussion

Before even getting into the significance of the different hypothesis tests run on the flux data for the star (KIC 8462852), when we look at the first graphical representation of the flux (of all orientations); which is plotted against the (Barycentric Julian Date) we see a clear decrease in the flux as time passes. It's easy to see that the results actually start at a maximum and bottom out at the last time input with a minimum. This already foreshadows the answer to the question, is the star increasing or decreasing in flux. Here though we can't take this at face value as the telescope is aligned for different orientations because each telescope orientation looks at a different part of the star so to make sure that our conclusions are accurate on whether the flux increases or decreases, we did this line graph representation for all the orientations done by the telescope.

From the 0th, 1st, 2nd and 3rd orientation we get a visible negative slope as time passes for all the scatter plots, so this does show that for all orientations that we always have a negative flux for the star, but this is only looking at it from a visual standpoint. To further ground and solidify our results we ran normal tests on all orientations for the flux and saw that our data for the orientations is distributed randomly so that we could use the p-value to set the null hypotheses for each so that we can either reject or accept them as per the p-value. For all the orientations of flux (looked at individually) the data was distributed randomly so the p-value was calculated to test against the null hypothesis of; the flux of the star is not decreasing and with an alternative hypothesis of, the star's flux decreasing. For all orientations, the null hypothesis was rejected because the p-value was found to be less than a 5 percent confidence level. The p-value was obtained by a one sample t- test as the logic that was used was that the first data point should be the mean because if the null hypothesis was correct (and there is no decrease in the flux) the value of the flux should not stray away from the initial value, but it does, very significantly. In addition to this regression analysis was used and the model with the highest R squared value turned out to be a regression model with degree 8 (R squared: 0.9555). This gave us a lot of confidence to reject the null hypothesis at 3 levels of interpretation: visual (from looking at the visual trend in the plots), numerical (from calculating and analyzing the p-value from normal distributed sets of data) and lastly through regression analysis.

There are some problems with our data collection, especially when looking at the data sets collected for fluxes at different orientations. To look further into this the orientations were analyzed purely on a categorical level, without even considering the fluxes. The frequency table, fig-, shows all the data points collected at different orientations. A one-way chi-squared test was used to find out whether these collected data points for different fluxes were random and the p-value says otherwise (less than 5% confidence) as it indicates that the data points taken for each orientation are not random. This is the first significant source of error in this data. The flux data points for different orientations should be random but even visually from fig- it can be seen that

52 flux data points were taken for the third orientation but only 11 for zero. That shows a heavy bias for our data because now it poses the question that if we had a more normally distributed set of data points for each orientation the results and the conclusion might have been different or would not lead to the same conclusion. For example, having 52 data points to look at and interpret for the third orientation would yield a much more reliable conclusion than a conclusion derived from only 11 data points (this is the 0th orientation).

Another source of error in our data is found when analyzing the values of uncertainty in the flux (uncertainty in y). The idea that the uncertainty does not change significantly for different flux (the null hypothesis) was rejected because the one sample t-test that compared the initial value of uncertainty to the entire uncertainty column, this was the same logic used to tell whether flux was decreasing, was found to be significantly different having a p-value of less than 0.05. Typically, the uncertainty in y should be constant (and if not close to being) which is exactly what was being tested because the same equipment and procedure is being used to find all the flux points (the only changing factor would be the orientation) but the uncertainty here is in y.

Conclusion

All the analyses conducted on the star leads to ultimately the same conclusion; The flux of the star KIC 846252 is significantly decreasing and this conclusion is reached by not only looking at the flux as time passes on its own but also further breaking it down to looking at it at 4 levels of orientation for the telescope, so that we can increase the reliability of our results but the introduction of different orientations should have been improved upon by randomizing the flux data points that were collected at different levels of orientation. The reason why this is emphasized is because both the uncertainty and the bias in more data points for certain orientations are significant sources of error, but the involvement of regression modeling does make the alternate hypothesis stronger. The star's reduction in flux on all fronts of the analysis does shed light towards more exploration and research of KIC 846252.

References

NASA. “Data products - TESS Science Support Center.” HEASARC, https://heasarc.gsfc.nasa.gov/docs/tess/data-products.html#full-frame-images. Accessed 11 April 2022.

STARR, MICHELLE. “The Mysterious 'Alien Megastructure' Star Is Not Alone, Astronomers Discover.” ScienceAlert, 20 January 2021, https://www.sciencealert.com/boyajian-s-star-isn-t-alone-astronomers-have-just-identified-a-binary-companion. Accessed 11 April 2022.

T. Montet, Benjamin, and Joshua D. Simon. “KIC 8462852 FADED THROUGHOUT THE KEPLER MISSION.” The Astrophysical Journal Letters [Pasadena], 20 October 2016, p. 13, https://iopscience.iop.org/article/10.3847/2041-8205/830/2/L39/pdf. Accessed 11 April 2022.