

# A Decision Tree for Classifying High Mass Pulsating Stars

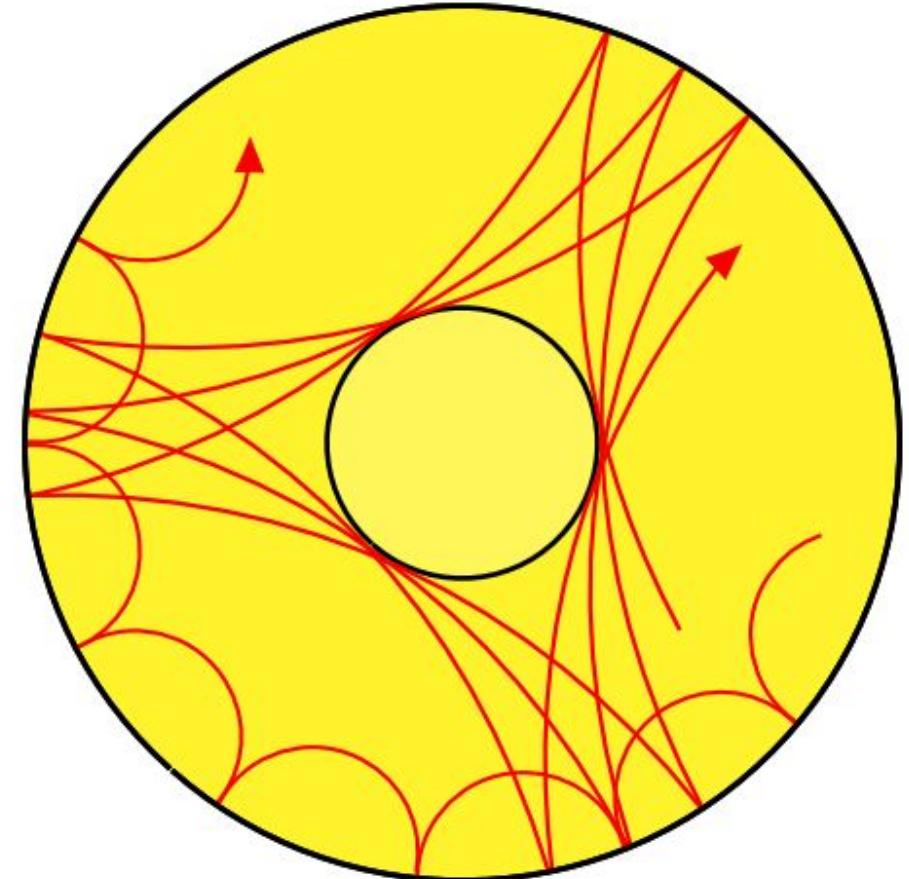
Mitchell Yenawine

Iowa State University

April 21, 2023

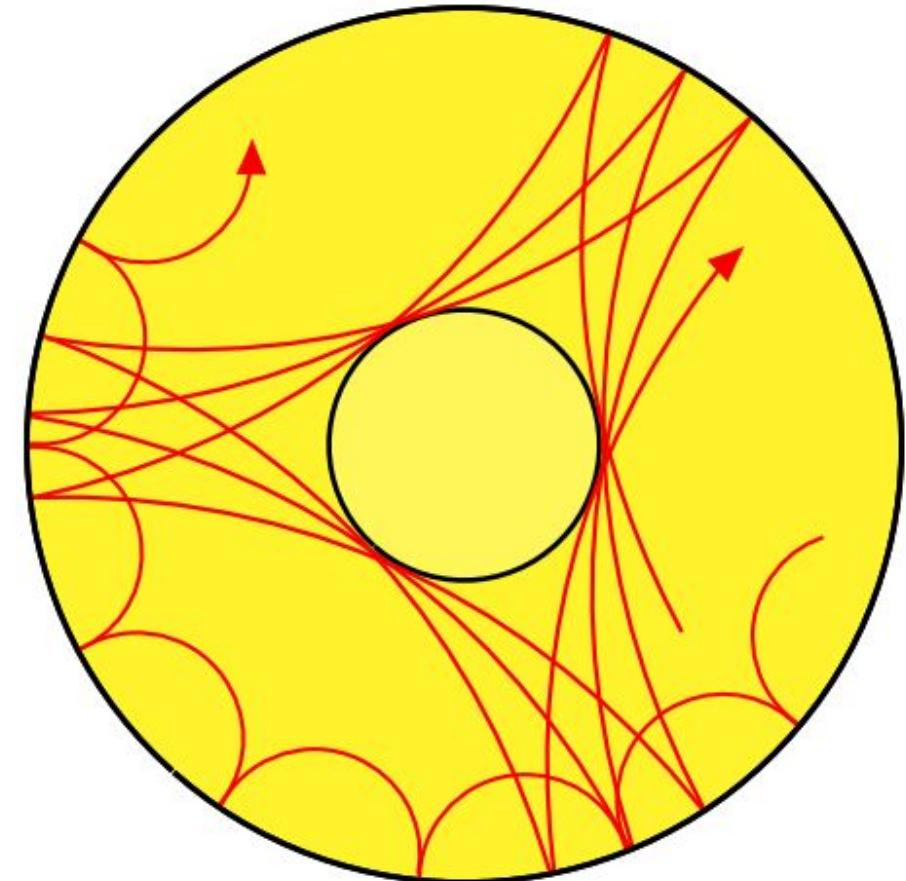
# Background: Asteroseismology

- Stars might pulsate if they have appropriate mass, radius, chemical composition, temperature, and age



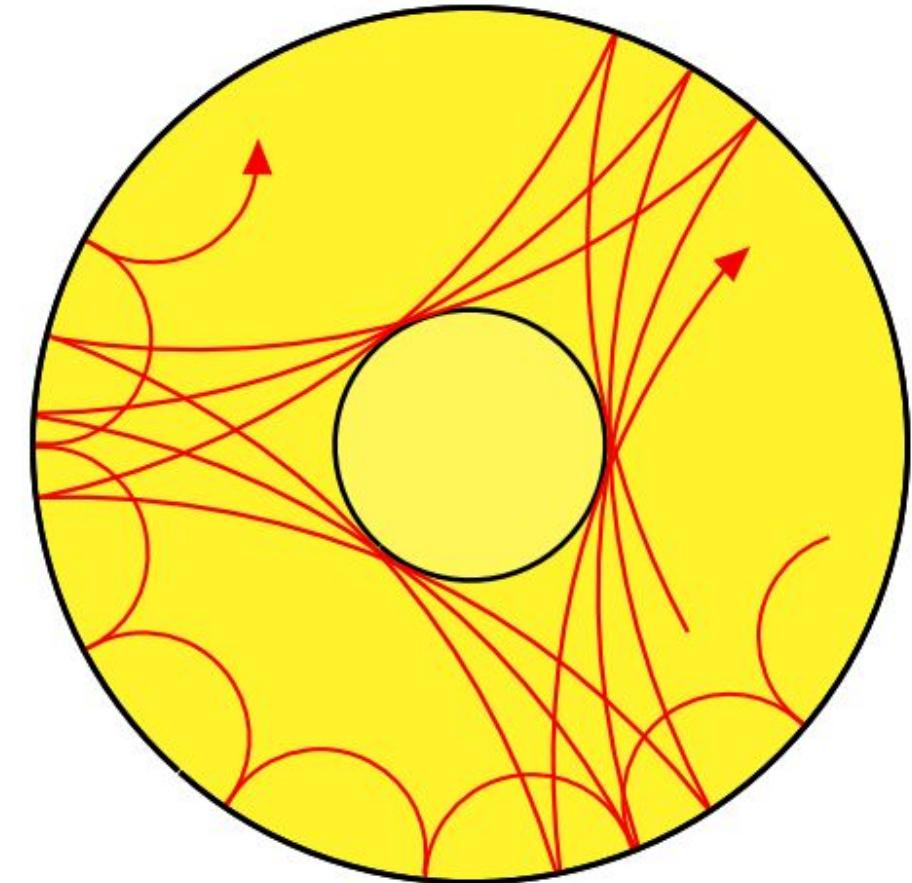
# Background: Asteroseismology

- Stars might pulsate if they have appropriate mass, radius, chemical composition, temperature, and age
- Driven by local instabilities that are then corrected for by either pressure or gravity:
  - convection, Kappa mechanism, tidal forces

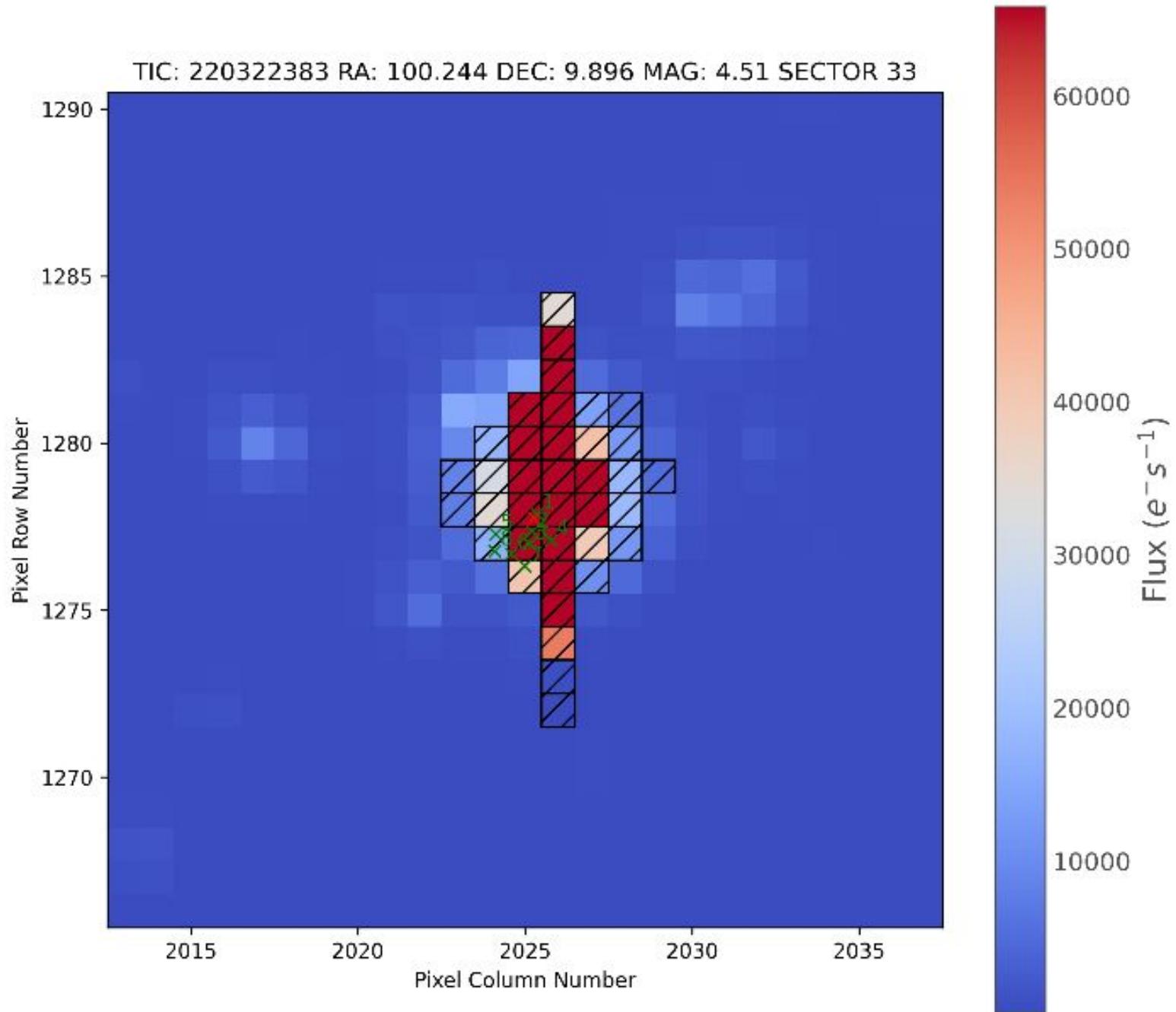


# Background: Asteroseismology

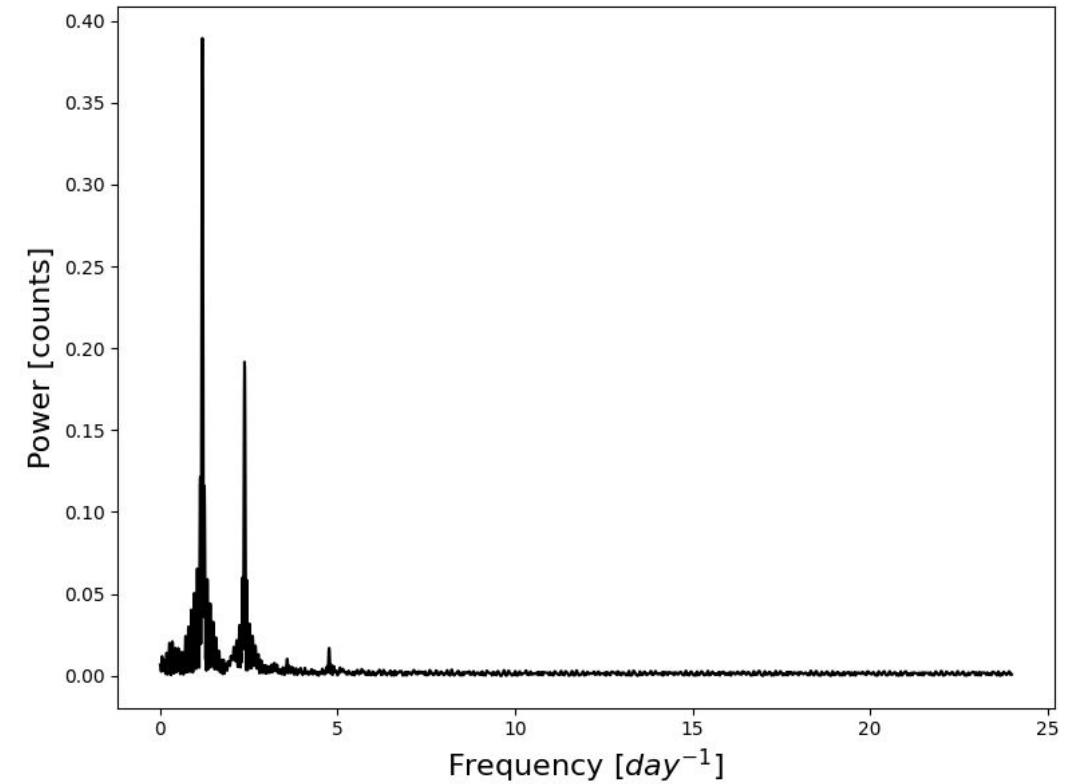
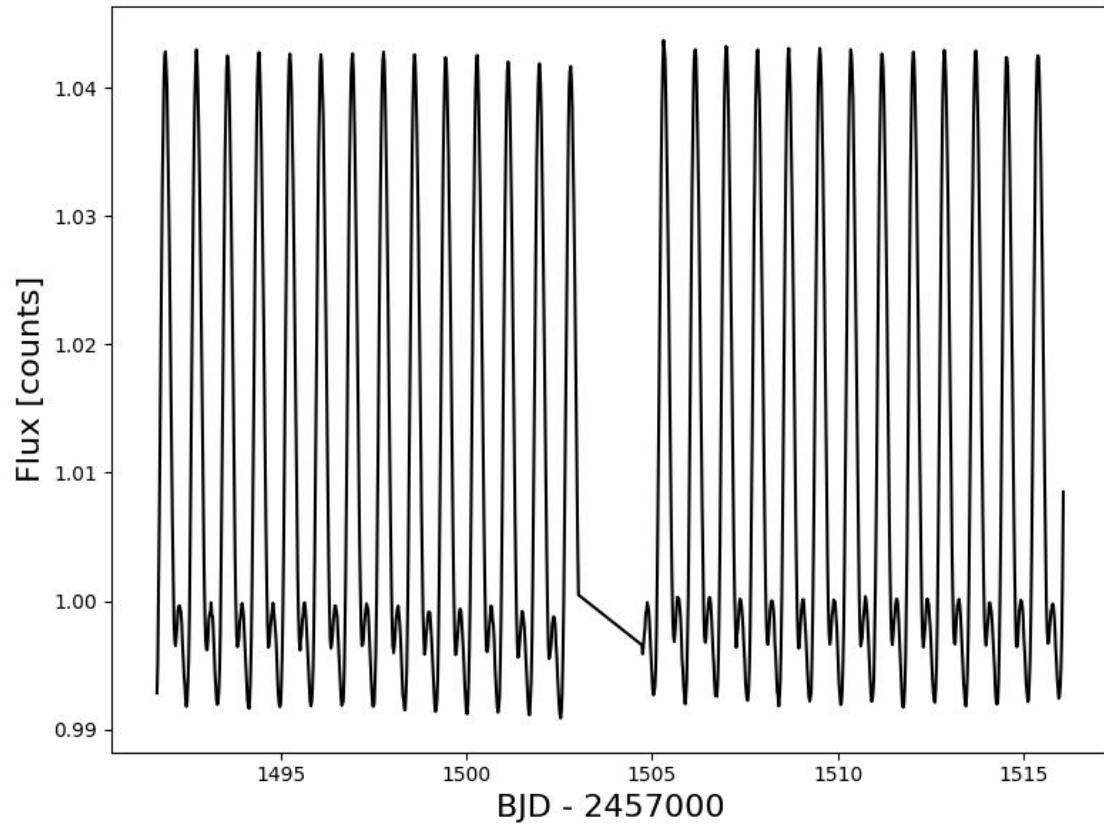
- Stars might pulsate if they have appropriate mass, radius, chemical composition, temperature, and age
- Driven by local instabilities that are then corrected for by either pressure or gravity:
  - convection, Kappa mechanism, tidal forces
- Probe the interiors of the stars bringing information about the interior to the surface



# Quick Example

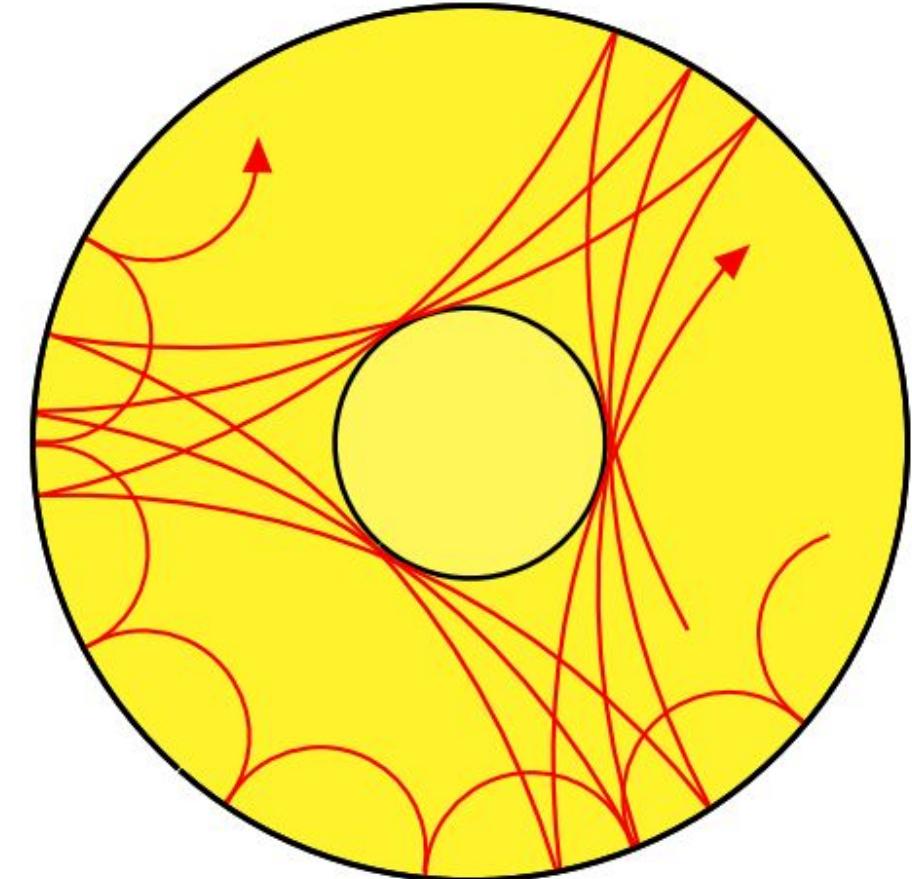


# Quick Example



# Background: Asteroseismology

- Stars might pulsate if they have appropriate mass, radius, chemical composition, temperature, and age
- Driven by local instabilities that are then corrected for by either pressure or gravity:
  - convection, Kappa mechanism, tidal forces
- Probe the interiors of the stars bringing information about the interior to the surface
- Oscillations and pulsations come in MANY different forms



# Background: Asteroseismology

American Association of Variable Star Observers (aavso)  
variability classifications:

- Eclipsing: 23
- Rotation: 18
- Pulsating: 60
- Eruptive: 34
- Cataclysmic: 47
- X-Ray: 11
- Other: 10



Source: <https://www.aavso.org>

# Background: Asteroseismology

- $\alpha^2$  Canum Venaticorum (ACV) - Main Sequence B and A stars with strong magnetic fields
  - Periods in the range of 0.5 to 160+ days
- $\beta$  Cephei (BCEP) - Main Sequence O and B stars with p and g modes pulsations.
  - Periods in the range of 0.1 to 0.6 days
- Slowly Pulsating B Stars (SPB) - B type stars with p and g mode pulsations.
  - Multiple periods may be present with ranges from 0.4 to 5 days

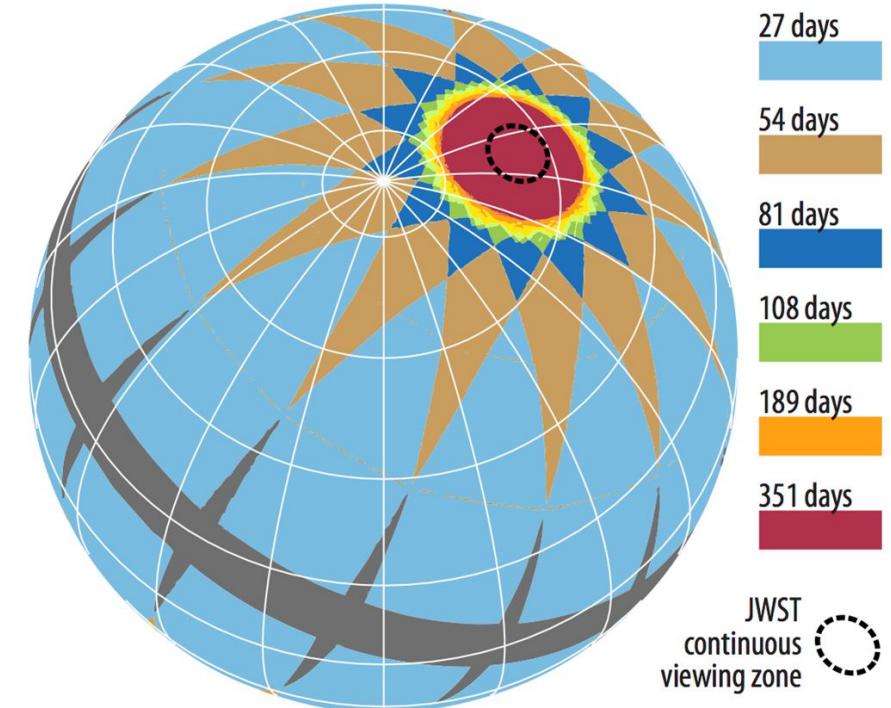


Source: <https://www.aavso.org>

# Transiting Exoplanet Survey Satellite (TESS)

- (Nearly) Full sky survey in both North and South hemisphere

TESS 2-year sky coverage map

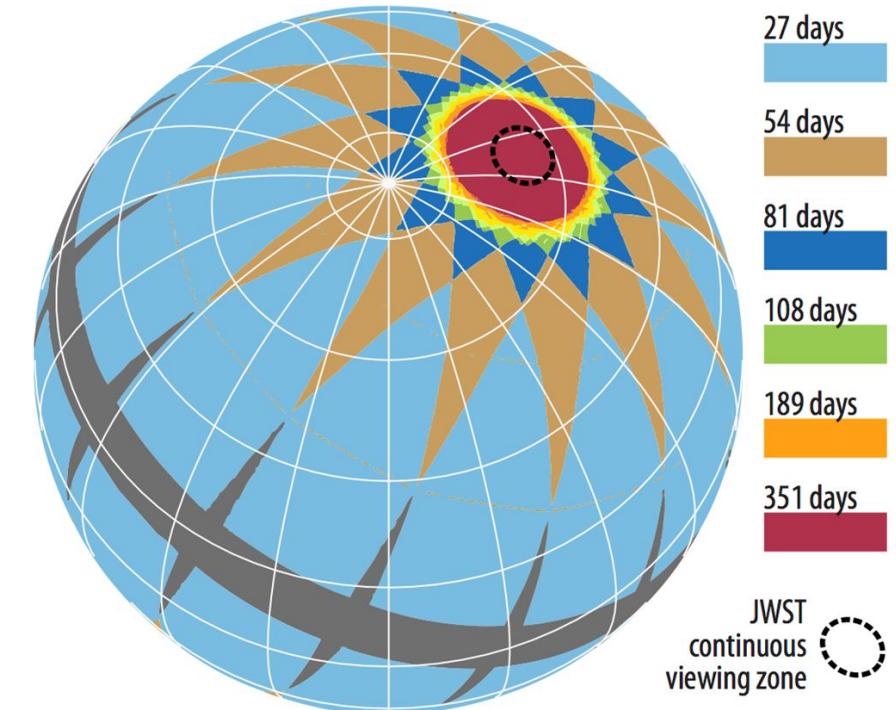


Source: <https://tess.mit.edu>

# Transiting Exoplanet Survey Satellite (TESS)

- (Nearly) Full sky survey in both North and South hemisphere
- Observes  $24^{\circ} \times 90^{\circ}$  sectors of the sky for 27.4 days
- Some regions may overlap between sectors leading to multiple 27.4 day observation periods

TESS 2-year sky coverage map

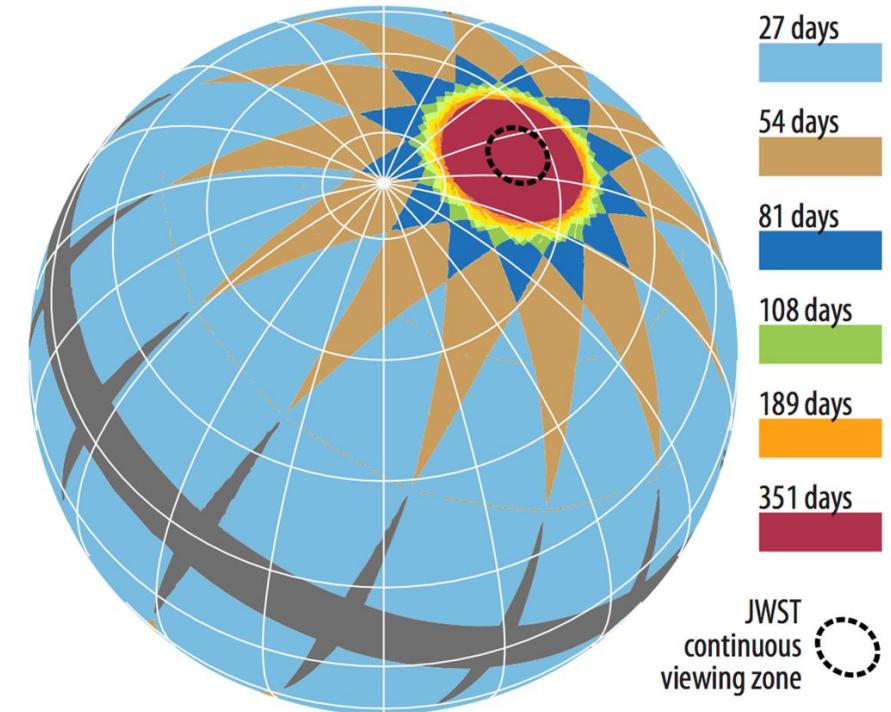


Source: <https://tess.mit.edu>

# Transiting Exoplanet Survey Satellite (TESS)

- (Nearly) Full sky survey in both North and South hemisphere
- Observes  $24^{\circ} \times 90^{\circ}$  sectors of the sky for 27.4 days
- Some regions may overlap between sectors leading to multiple 27.4 day observation periods
- Very large pixels at  $21''$  per pixel
  - Typical ground based telescope might be  $0.4''$  per pixel
  - Need to be careful of contaminating flux from nearby stars

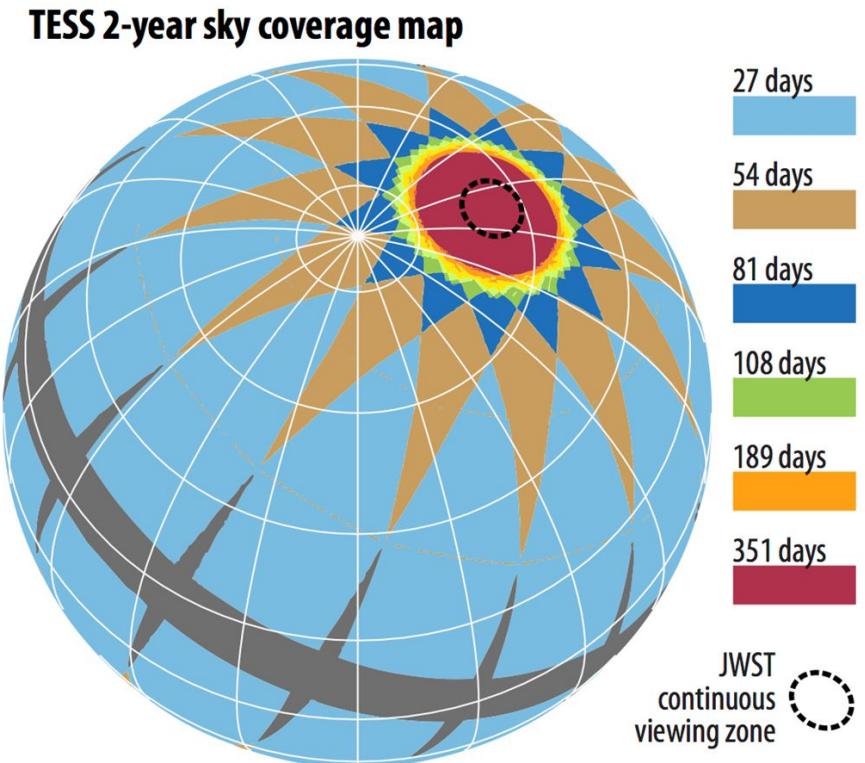
TESS 2-year sky coverage map



Source: <https://tess.mit.edu>

# Transiting Exoplanet Survey Satellite (TESS)

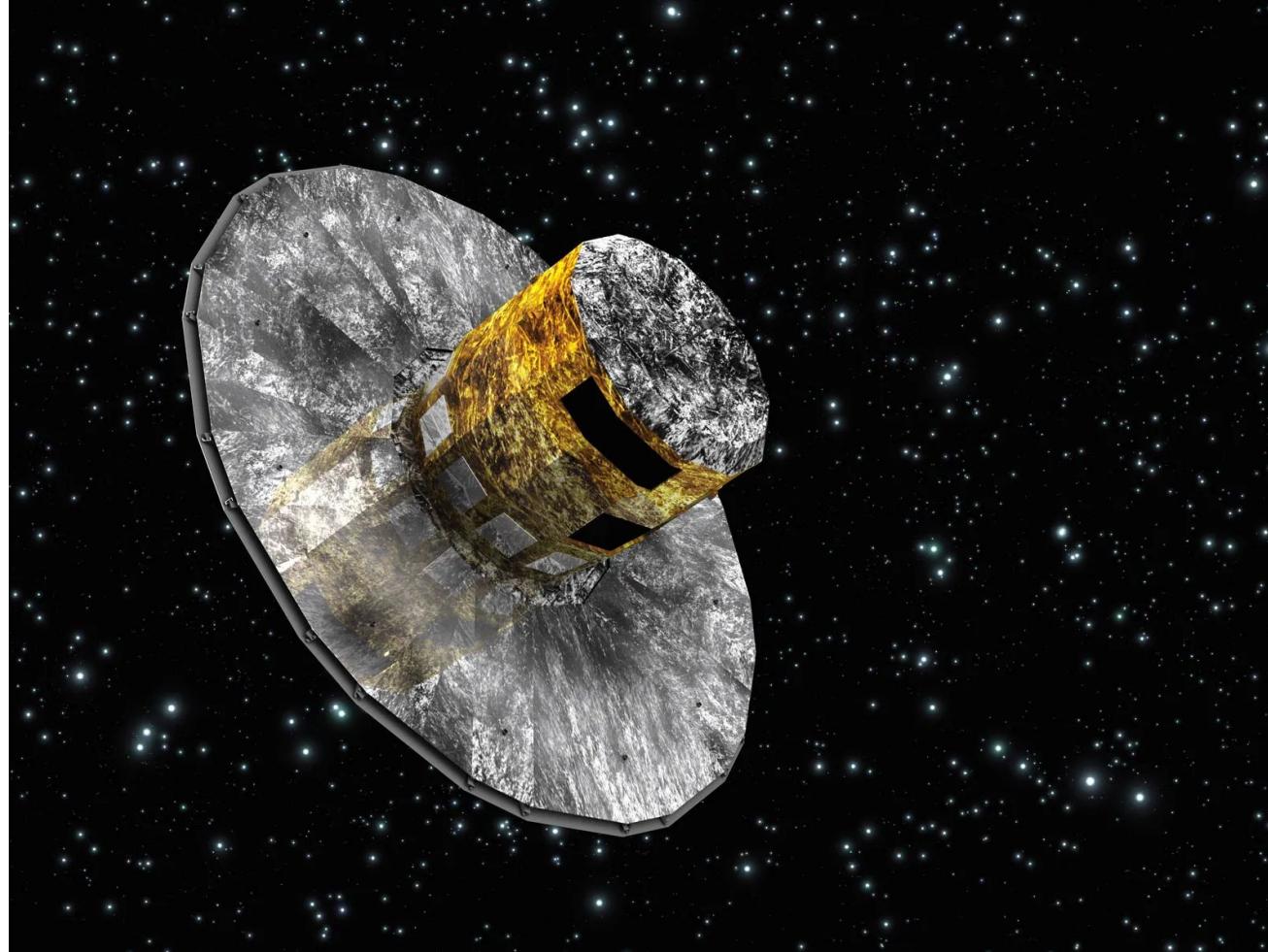
- (Nearly) Full sky survey in both North and South hemisphere
- Observes  $24^{\circ} \times 90^{\circ}$  sectors of the sky for 27.4 days
- Some regions may overlap between sectors leading to multiple 27.4 day observation periods
- Very large pixels at  $21''$  per pixel
  - Typical ground based telescope might be  $0.4''$  per pixel
  - Need to be careful of contaminating flux from nearby stars
- Full Frame Images (FFIs) of entire sky available at 30 minute cadence and select objects observed at 2 minute cadence



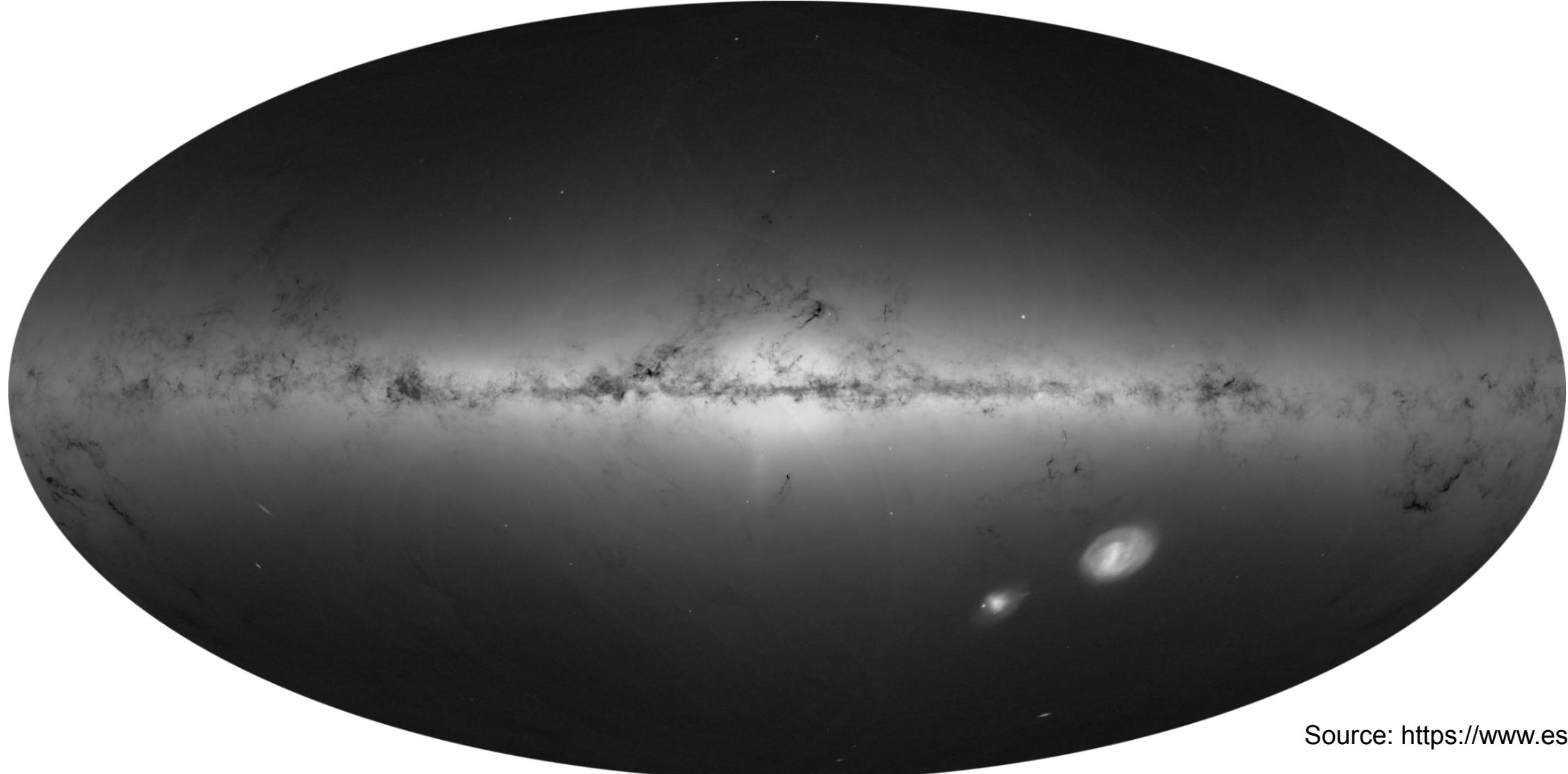
Source: <https://tess.mit.edu>

# Gaia Spacecraft

Observed the positions, distance, proper motion, and colors of many stars

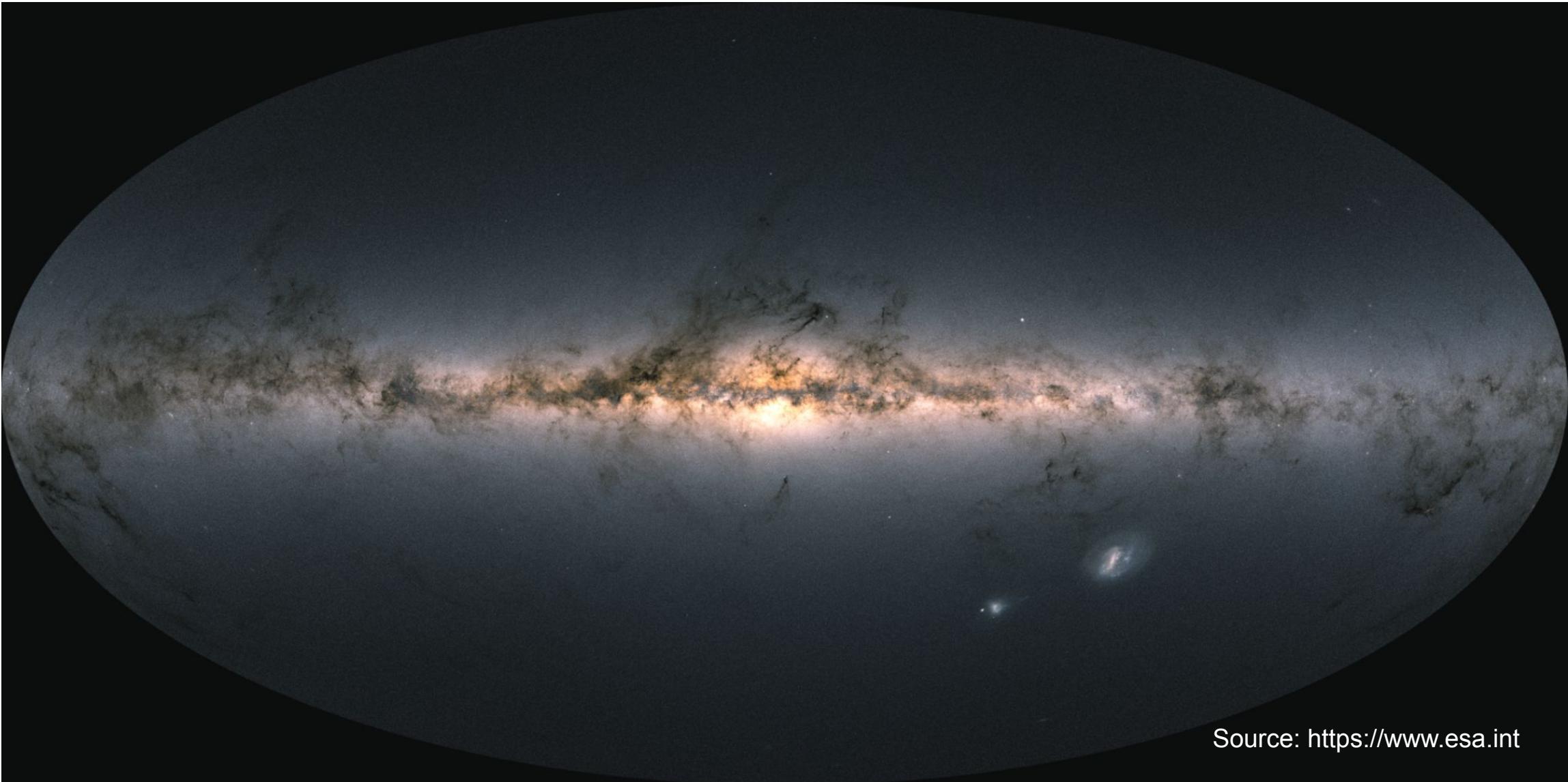


# Gaia Spacecraft



Source: <https://www.esa.int>

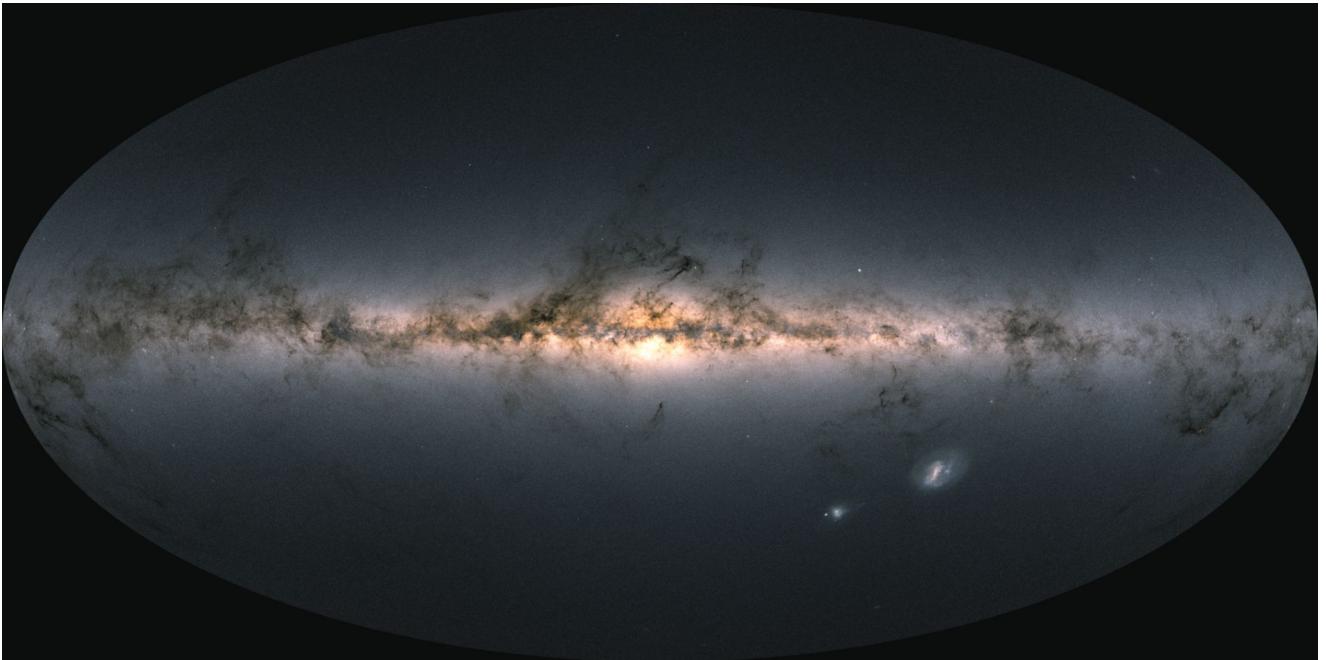
# Gaia Spacecraft



Source: <https://www.esa.int>

# Gaia Spacecraft

- Positions and brightness of ~1.8 billion stars
- Parallax (distance) and Colors of ~1.5 billion stars



Source: <https://www.esa.int>

# Targets

Stars already classified by AAVSO and observed by gaia and TESS:

- ACV - 417
- BCEP - 68
- SPB - 79

# Targets

Stars already classified by AAVSO and observed by gaia and TESS:

- ACV - 417
- BCEP - 68
- SPB - 79

After weeding out anomalies and reducing the number of ACV stars to roughly match the others:

- ACV - 71
- BCEP - 40
- SPB - 28

# From TESS and Gaia

- TESS:
  - Light Curves
  - After Lomb-Scargle or FFT: Periods, Power, Amplitudes
- Gaia:
  - Colors, Effective Temperatures

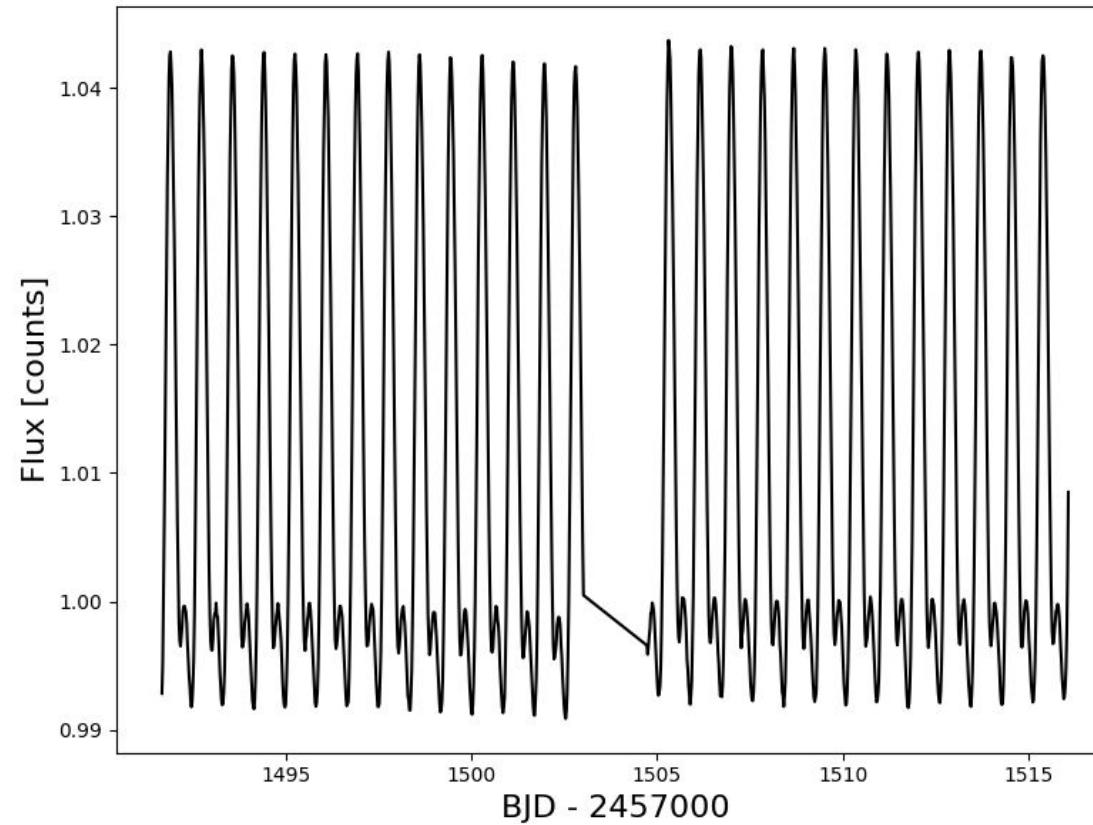
# From TESS and Gaia

- TESS:
  - Light Curves
  - After Lomb-Scargle or FFT : Periods, Power, Amplitudes
- Gaia:
  - Colors, Effective Temperatures

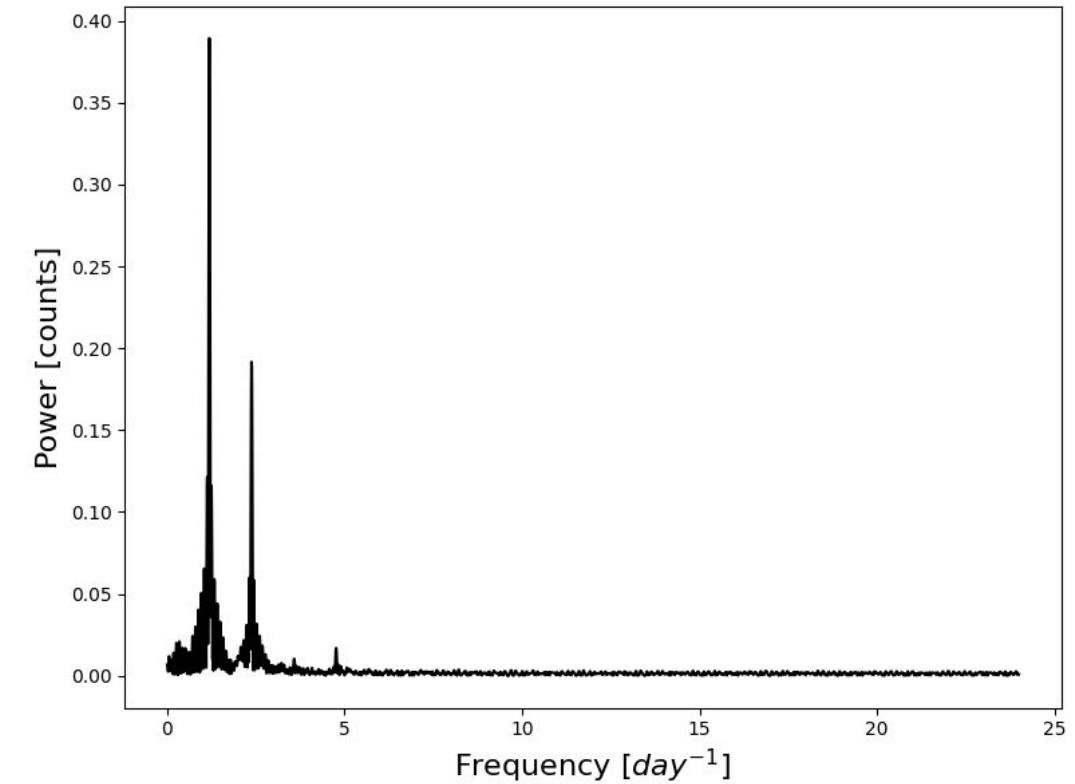
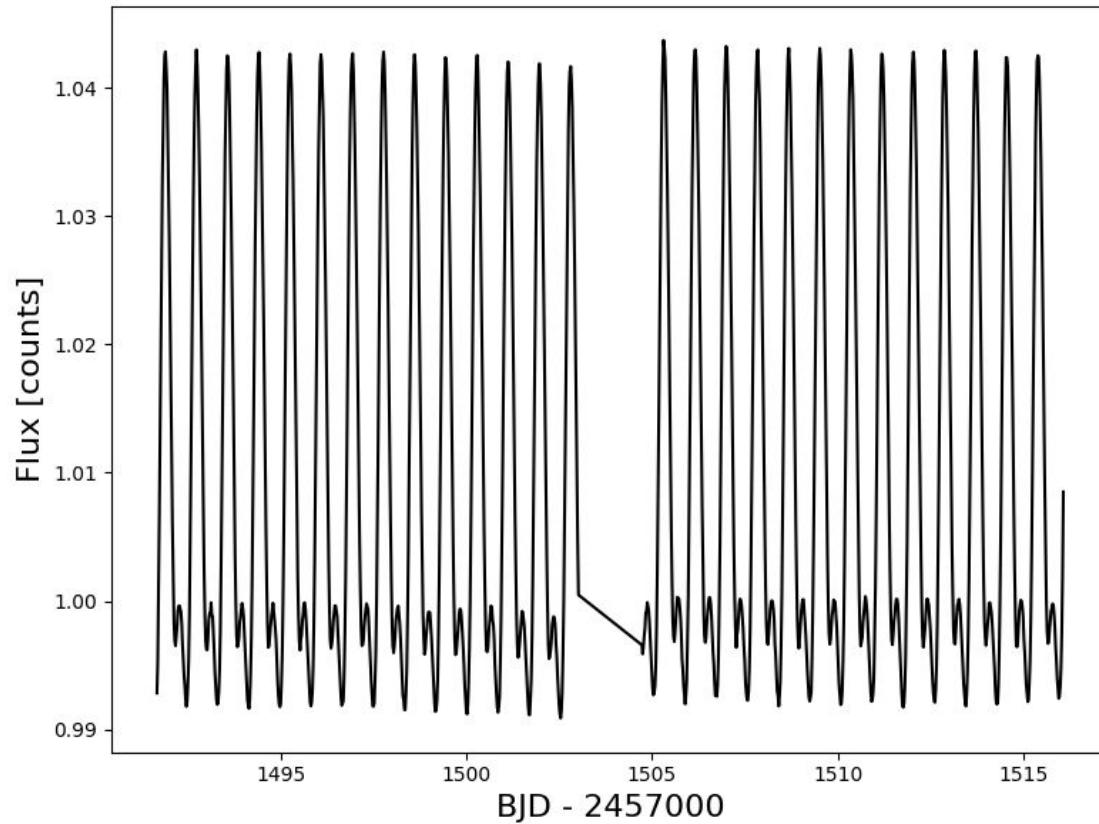
Take a bit more work....

Pulled straight from Gaia

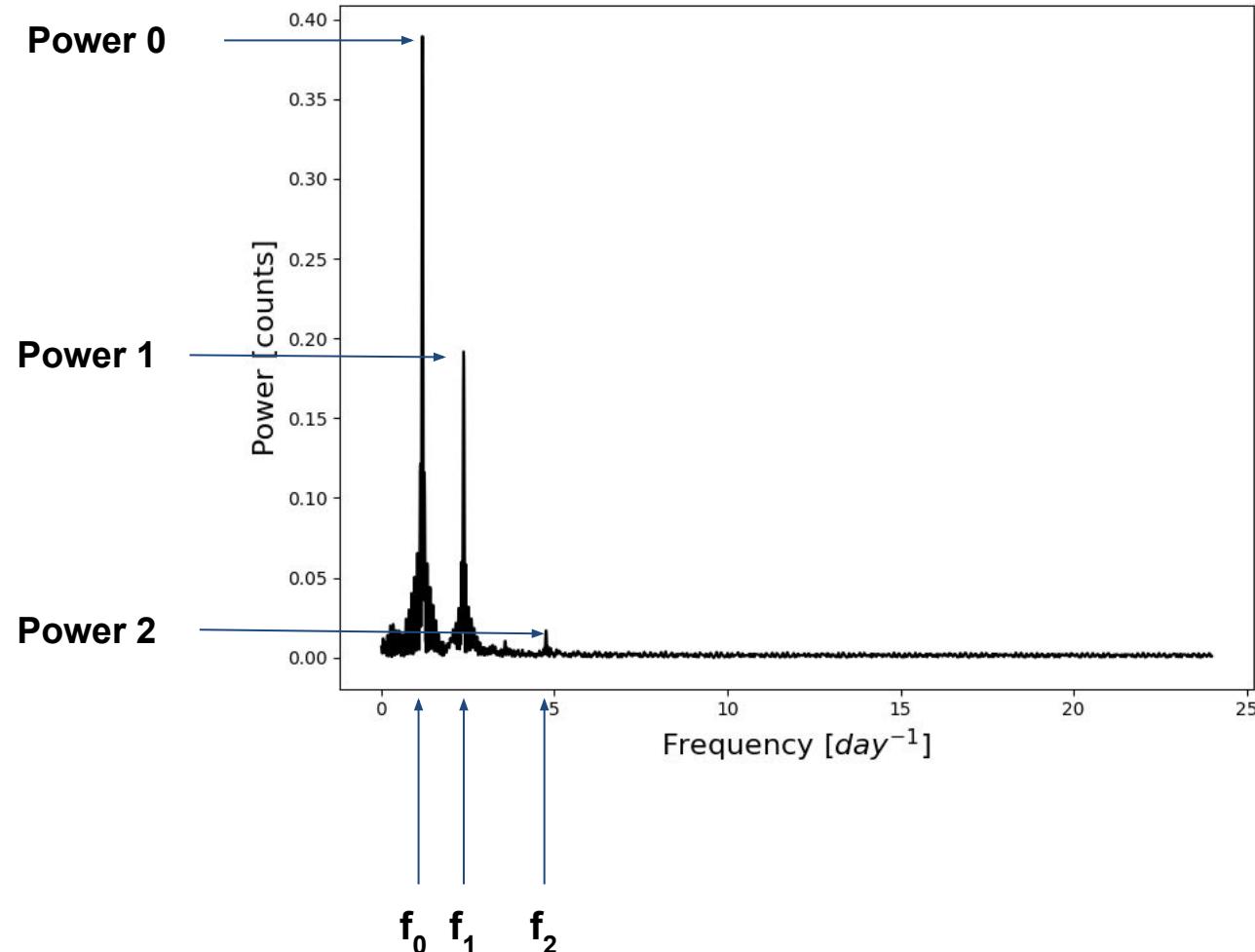
# Quick Example Again



# Quick Example Again



# Quick Example Again

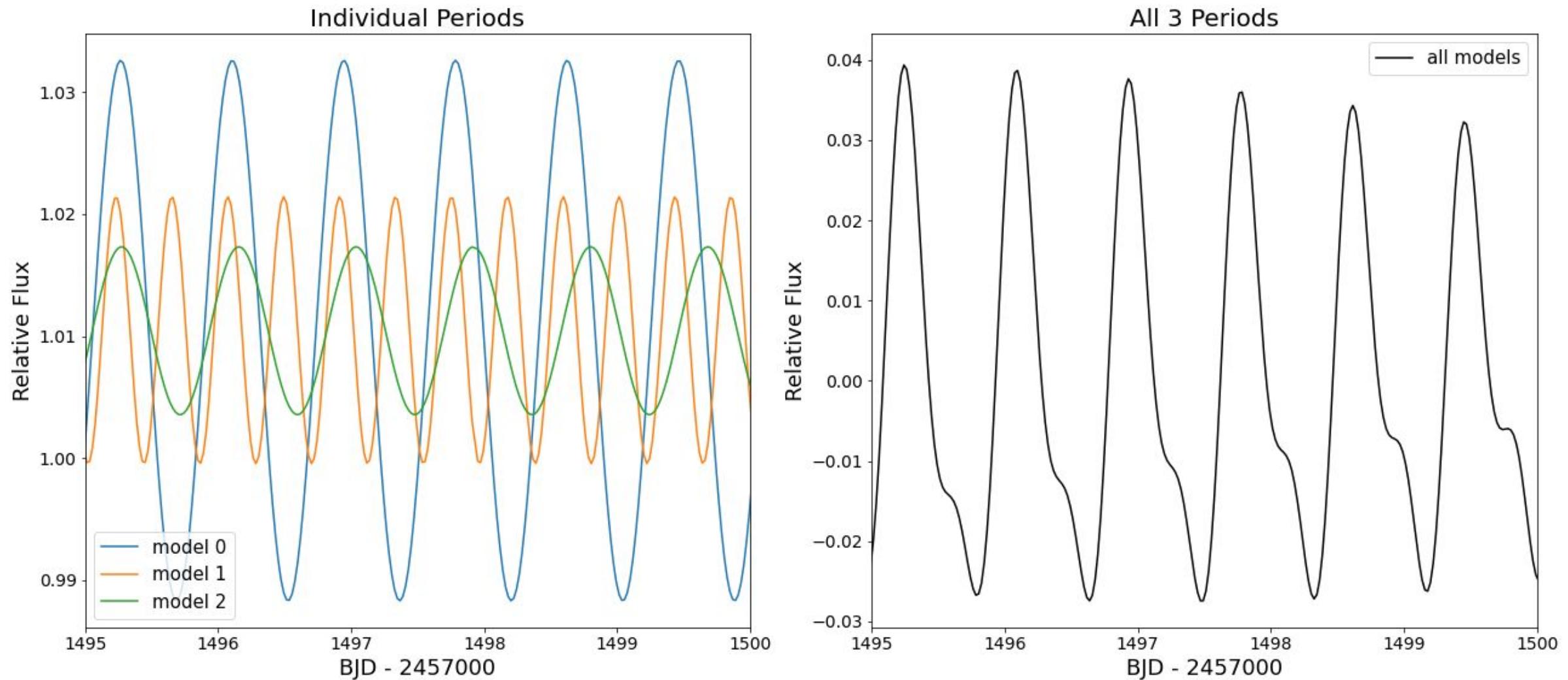


$$\text{Period } 0 = 1/f_0$$

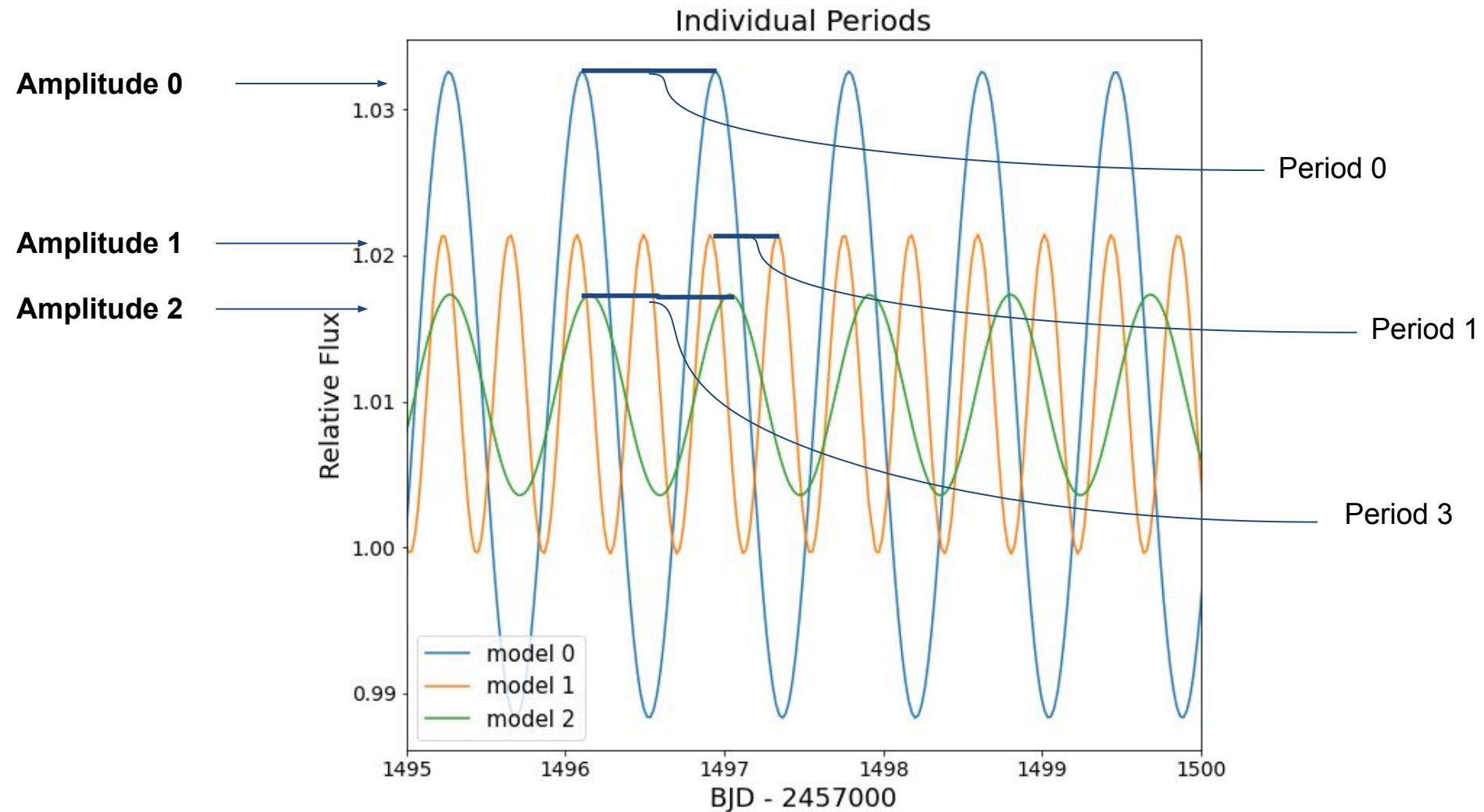
$$\text{Period } 1 = 1/f_1$$

$$\text{Period } 2 = 1/f_2$$

# Quick Example Again

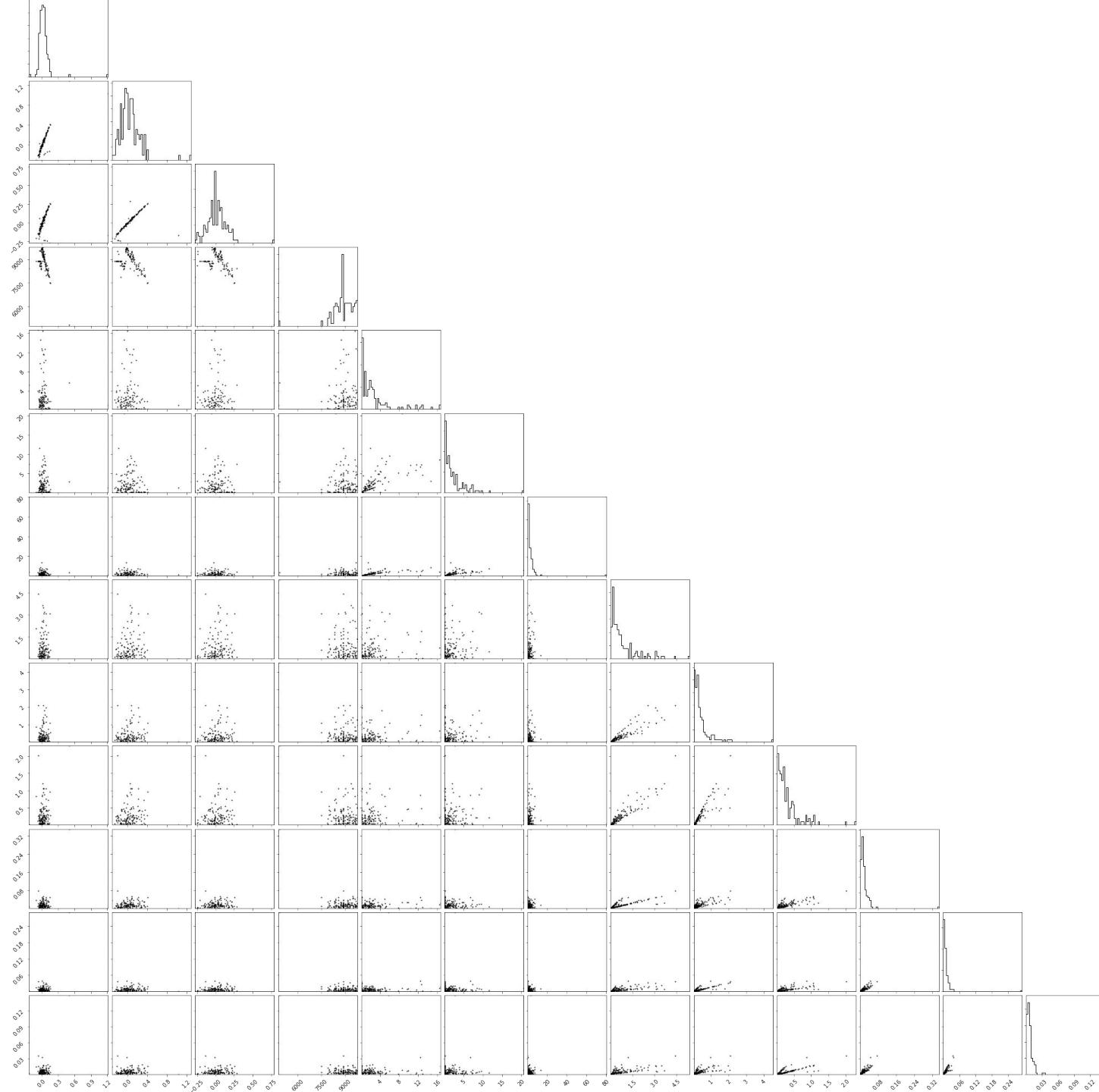


# Quick Example Again



# Features

- In total: 13
  - 3 colors: (B - V), (B - R), (V - R)
  - Effective Temperature: Teff
  - Periods, Powers, and Amplitudes of 3 strongest pulsations



# Features

- In total: 13
  - 3 colors: (B - V), (B - R), (V - R)
  - Effective Temperature: Teff
  - Periods, Powers, and Amplitudes of 3 strongest pulsations
- Other Possible Features?
  - frequency spacing, harmonics, number of strong pulsation modes, red noise parameters, signal-to-noise?

# Implementation: Decision Tree

- Was planning on modifying HW 3.....

# Implementation: Decision Tree in Scikit-learn

- Was planning on modifying HW 3.....
- Ran out of time due to Final Oral Exam/Presentation so used Scikit-learn

# Implementation: Decision Tree in Scikit-learn

- Hard part: getting the features
  - ~300 lines of python code running for 2 hours on pronto

# Implementation: Decision Tree in Scikit-learn

- Hard part: getting the features
  - ~300 lines of python code running for 2 hours on pronto
    - multiple times...

# Implementation: Decision Tree in Scikit-learn

- Hard part:

```
1 import numpy as np
2
3 feature_names = ['B-V', 'B-R', 'V-R', 'Teff',
4                   'period0', 'period1', 'period2',
5                   'power0', 'power1', 'power2',
6                   'amp0', 'amp1', 'amp2']
7
8 label_names = ['ACV', 'BCEP', 'SPB']
9
10
11 features = np.load('features.npy')
12 features = features[1:]
13 cls = np.load('classes.npy')
14
15 w = np.asarray([np.all(f == f) for f in features])
16 features = features[w]
17 cls = cls[w]
18
19 targets = np.zeros(len(cls))
20
21 for i, name in enumerate(label_names):
22     w = cls == name
23     print(name, sum(w))
24     targets[w] = i
25
```

ACV 71

BCEP 40

SPB 28

# Implementation: Decision Tree in Scikit-learn

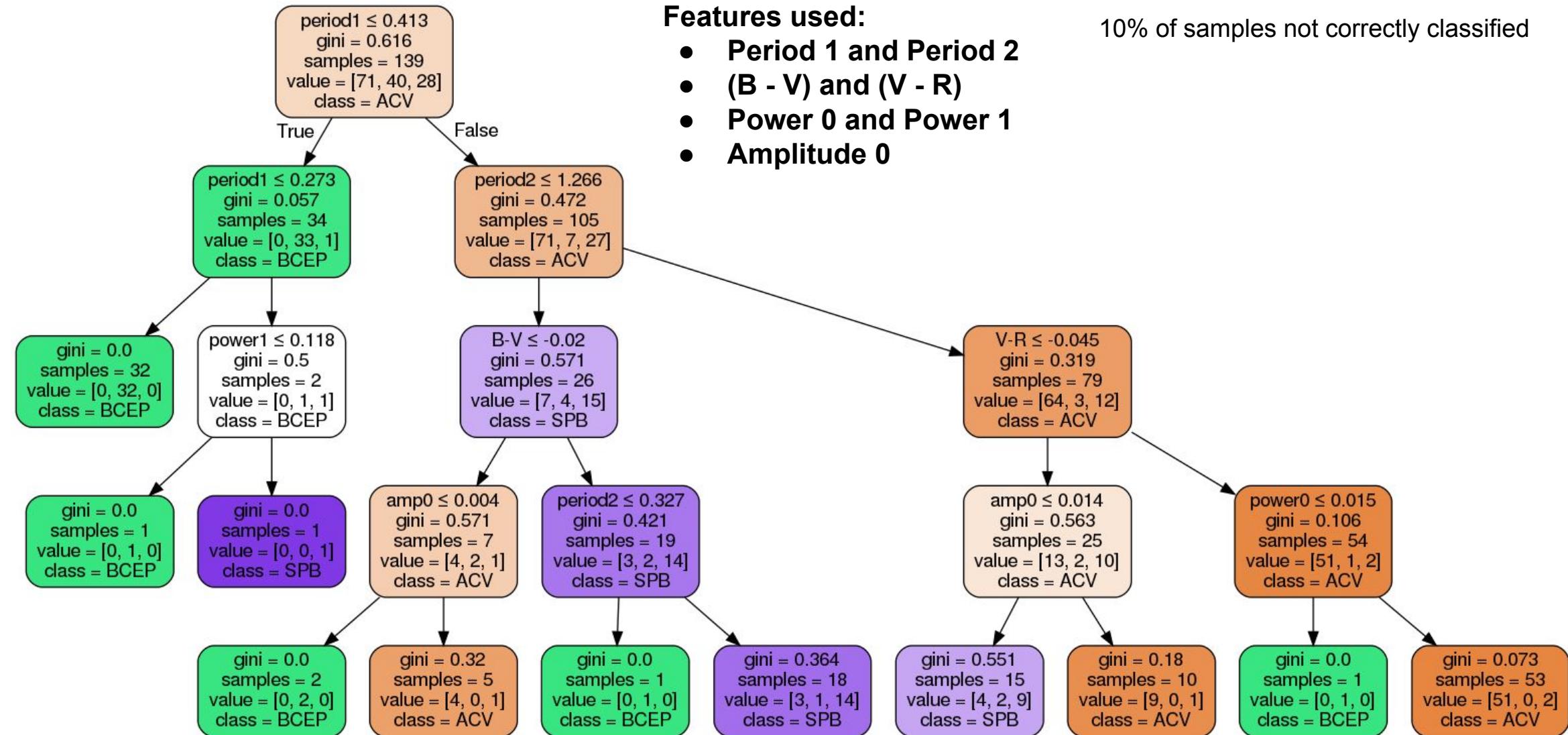
- Easy part:

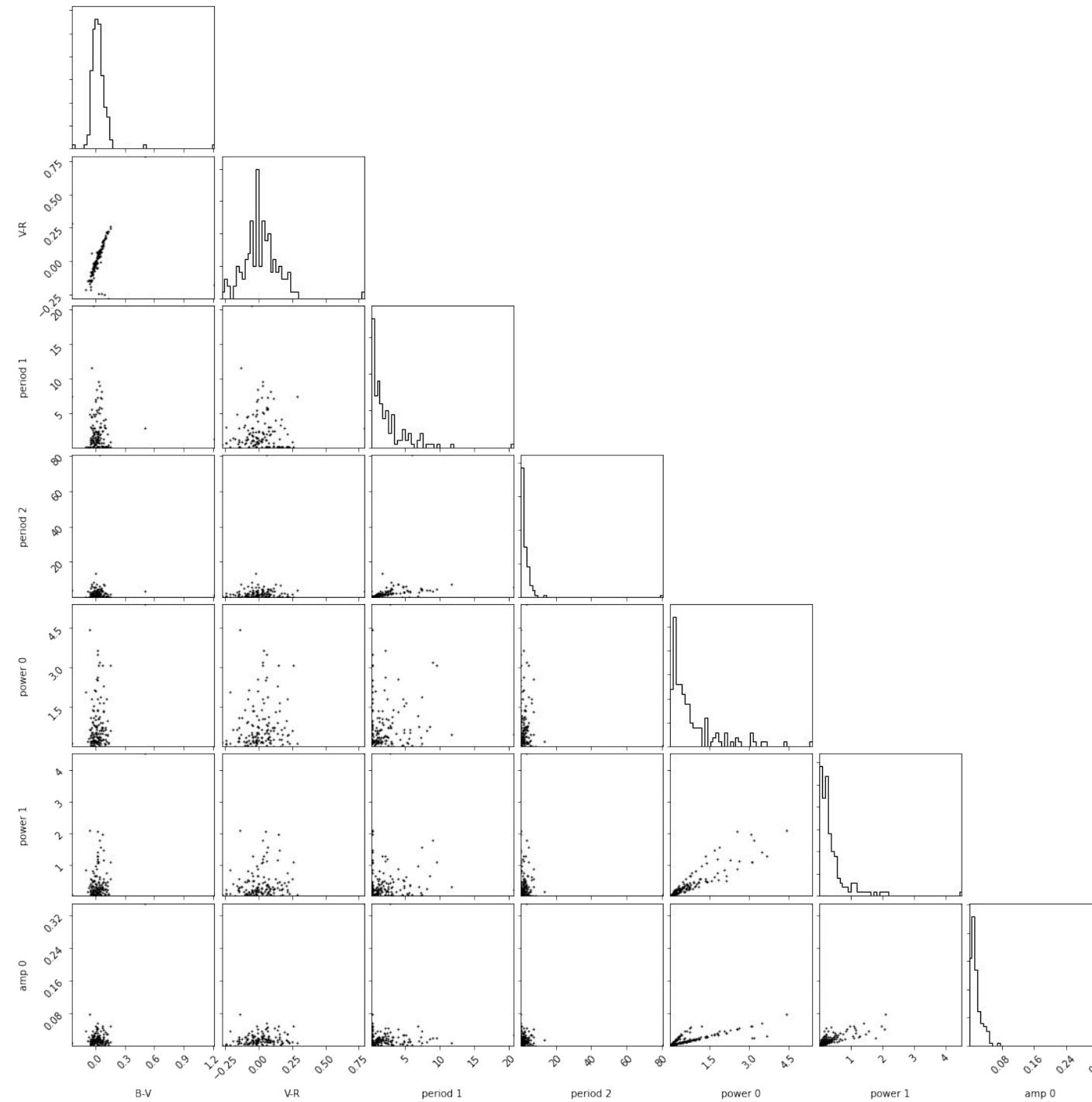
```
1 from sklearn import tree
2
3 num_samples = len(features)
4 bad_predictions = True
5 level_stop = 12
6 levels = 0
7
8 while bad_predictions and levels < level_stop:
9     levels += 1
10
11    decision_tree = tree.DecisionTreeClassifier(random_state=0, max_depth=levels)
12    decision_tree = decision_tree.fit(features, targets)
13    r = tree.export_text(decision_tree, feature_names=feature_names)
14
15    wrong_predictions = 0
16    for f, t in zip(features, targets):
17        if decision_tree.predict([f])[0] == t:
18            continue
19        else:
20            wrong_predictions += 1
21
22    print('With {} levels {} % of samples are wrongly predicted'.format(levels, round(100*wrong_predictions/num_samples, 2)))
23
24    fig, ax = plt.subplots(1, 1, figsize=(25,12))
25
26    _ = tree.plot_tree(decision_tree=decision_tree, ax=ax, feature_names=feature_names, fontsize=16)
27
28    plt.figure('levels_{}.png'.format(levels))
29
30    plt.close()
31
32    if wrong_predictions == 0:
33        bad_predictions = False
34
```

# Implementation: Decision Tree in Scikit-learn

```
With 1 levels 25.18 % of samples are wrongly predicted  
With 2 levels 19.42 % of samples are wrongly predicted  
With 3 levels 16.55 % of samples are wrongly predicted  
With 4 levels 10.07 % of samples are wrongly predicted  
With 5 levels 7.19 % of samples are wrongly predicted  
With 6 levels 2.16 % of samples are wrongly predicted  
With 7 levels 0.72 % of samples are wrongly predicted  
With 8 levels 0.0 % of samples are wrongly predicted
```

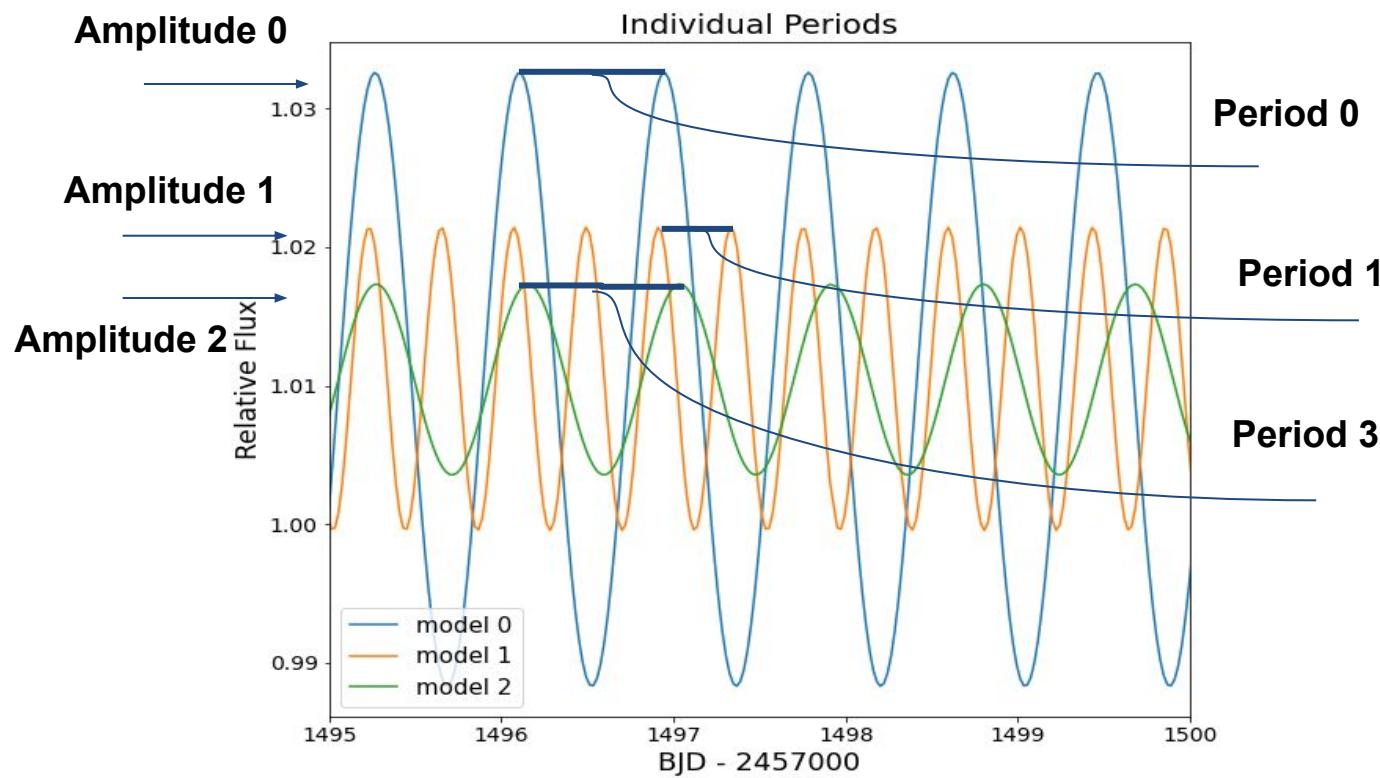
# The Decision Tree (4 levels)





# Interesting things from the Decision Trees

- By far the most important feature is the period of the second strongest pulsating mode **Period 1**
- The period of the strongest pulsation (**Period 0**) is not used at any level, but the amplitude and power are (**Amplitude 0 and Power 0**)
- B - V and V - R are the only colors needed
  - B - R can be made from these colors
- Effective Temperature not needed
  - Can be derived from colors B - V and V - R

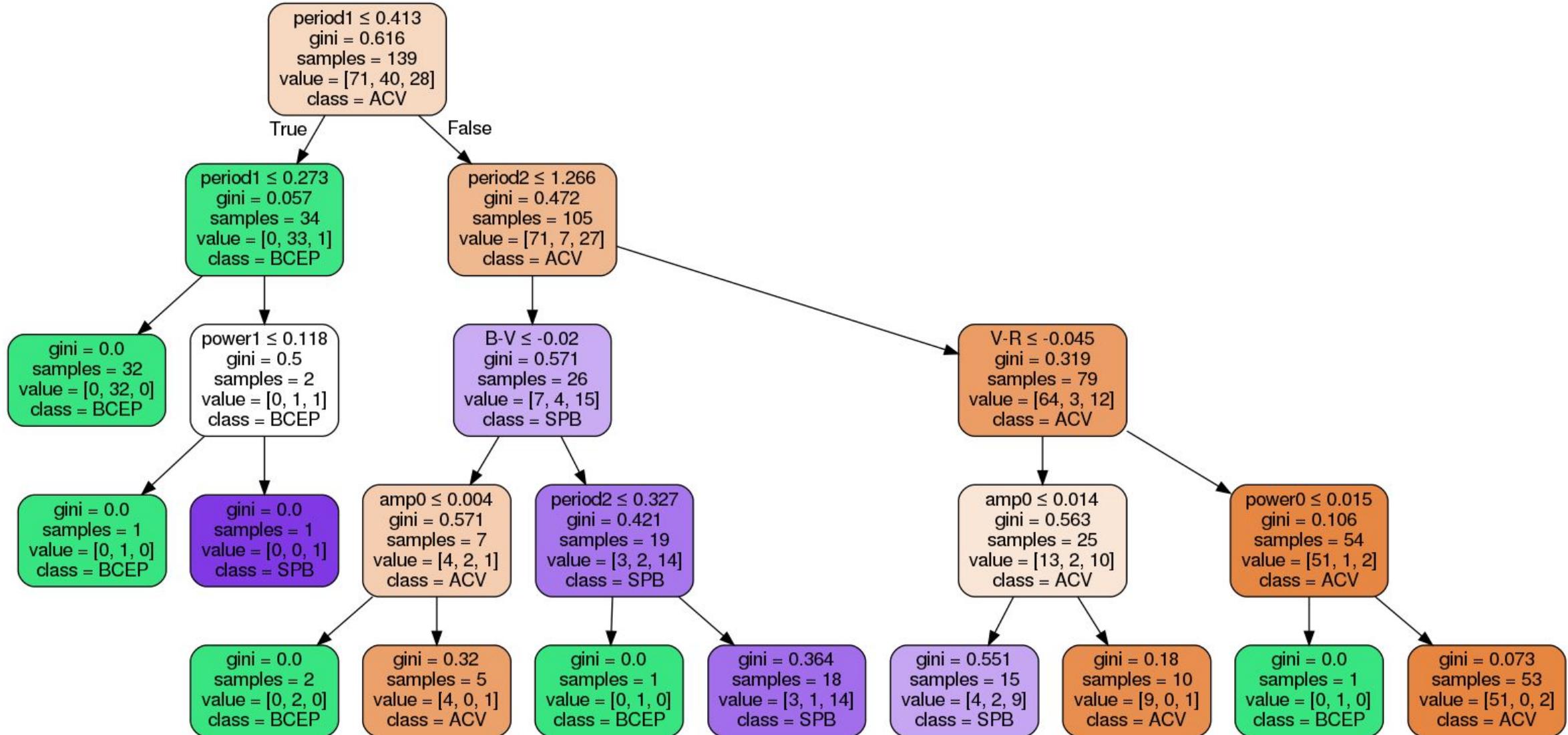


# Future work

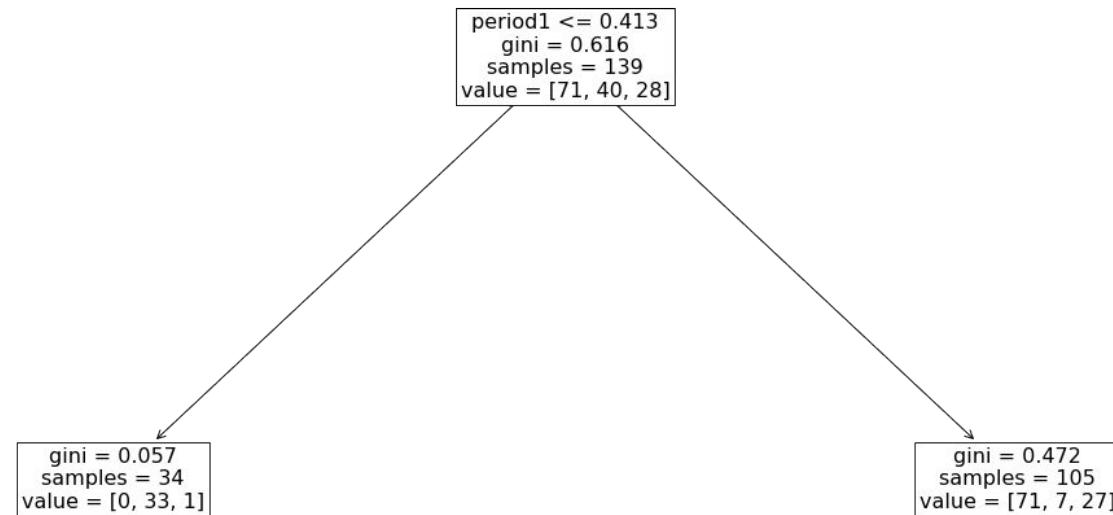
Tons:

- Add other pulsation classifications
  - Especially a ‘Null’ case
- Clean up data further
  - not all flux data are in the same units (counts vs relative flux)
  - Binary systems
- Add other TESS sectors to the data
  - Periods often longer than 27 days
- Check other features
  - frequency spacing, red noise, etc.

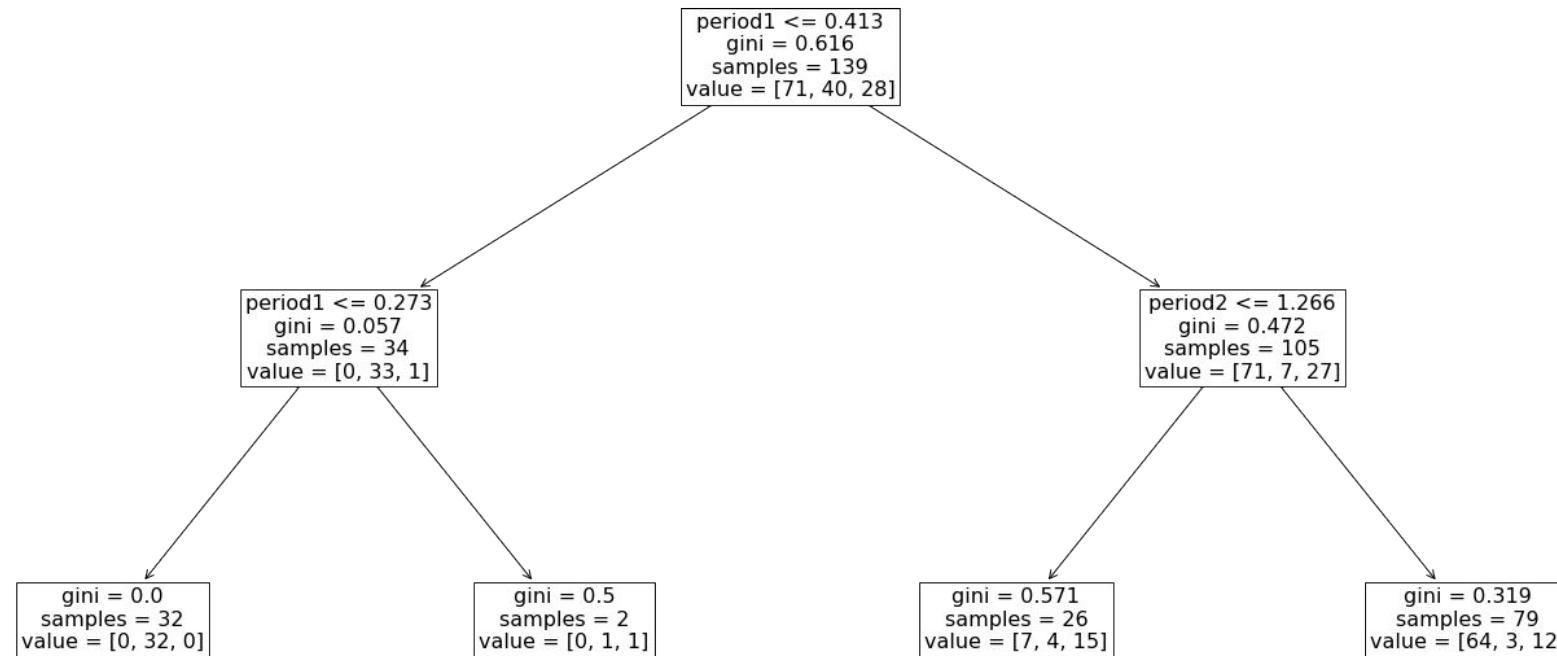
# Thanks for listening! Questions?



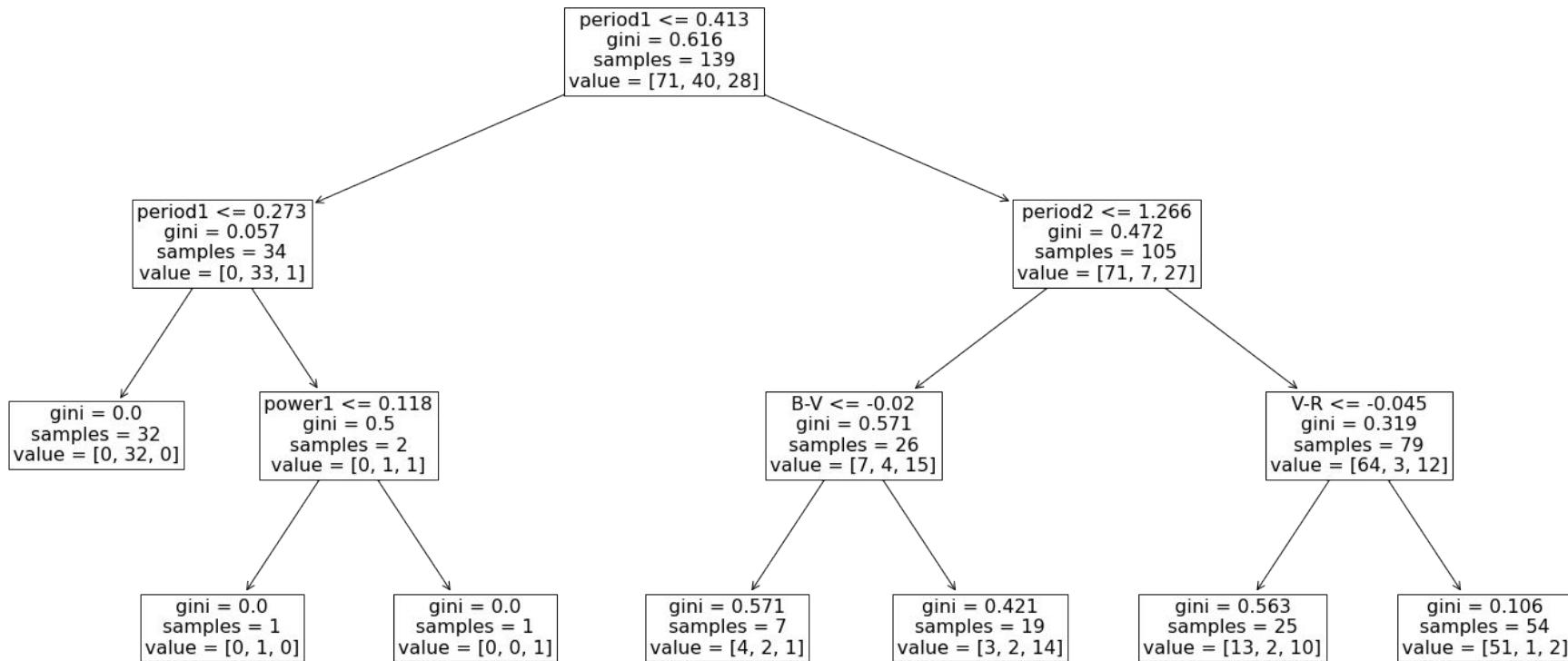
# Extra: Decision Tree with Different Levels 1



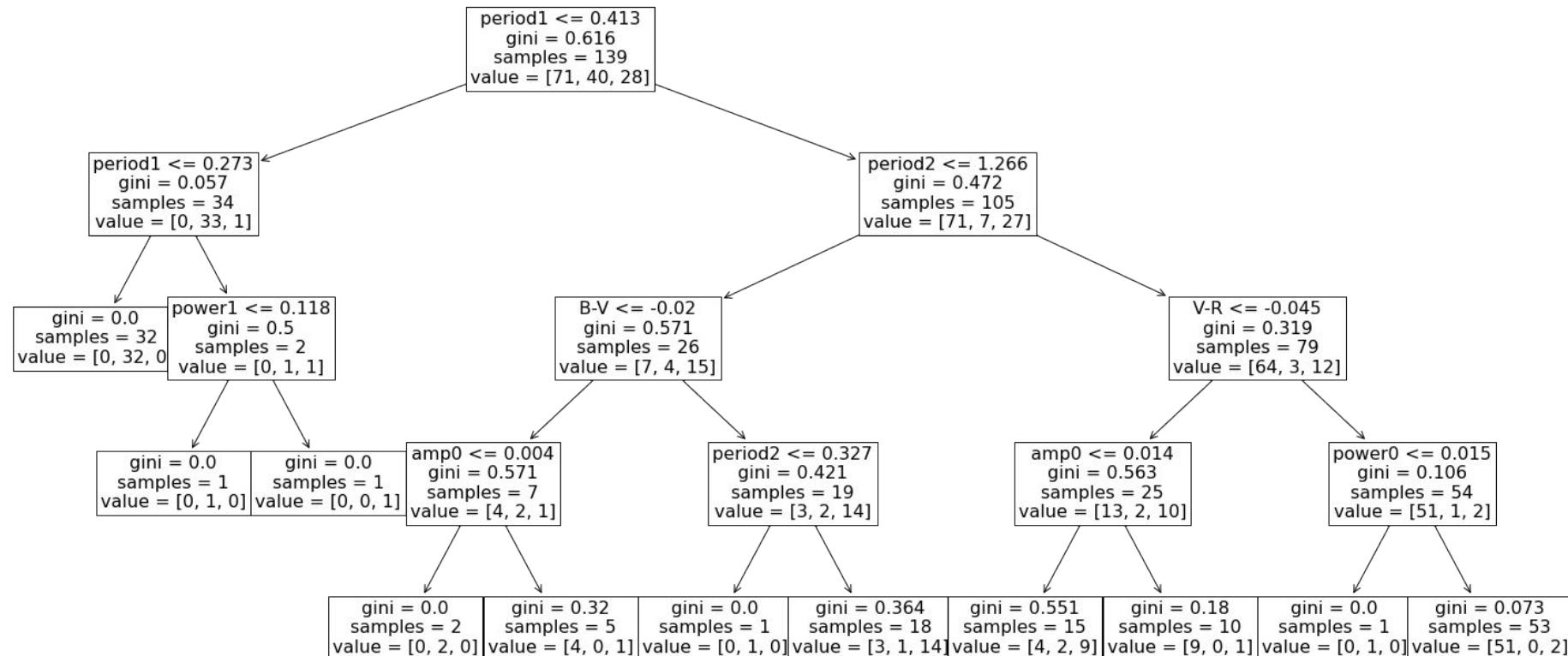
# Extra: Decision Tree with Different Levels 2



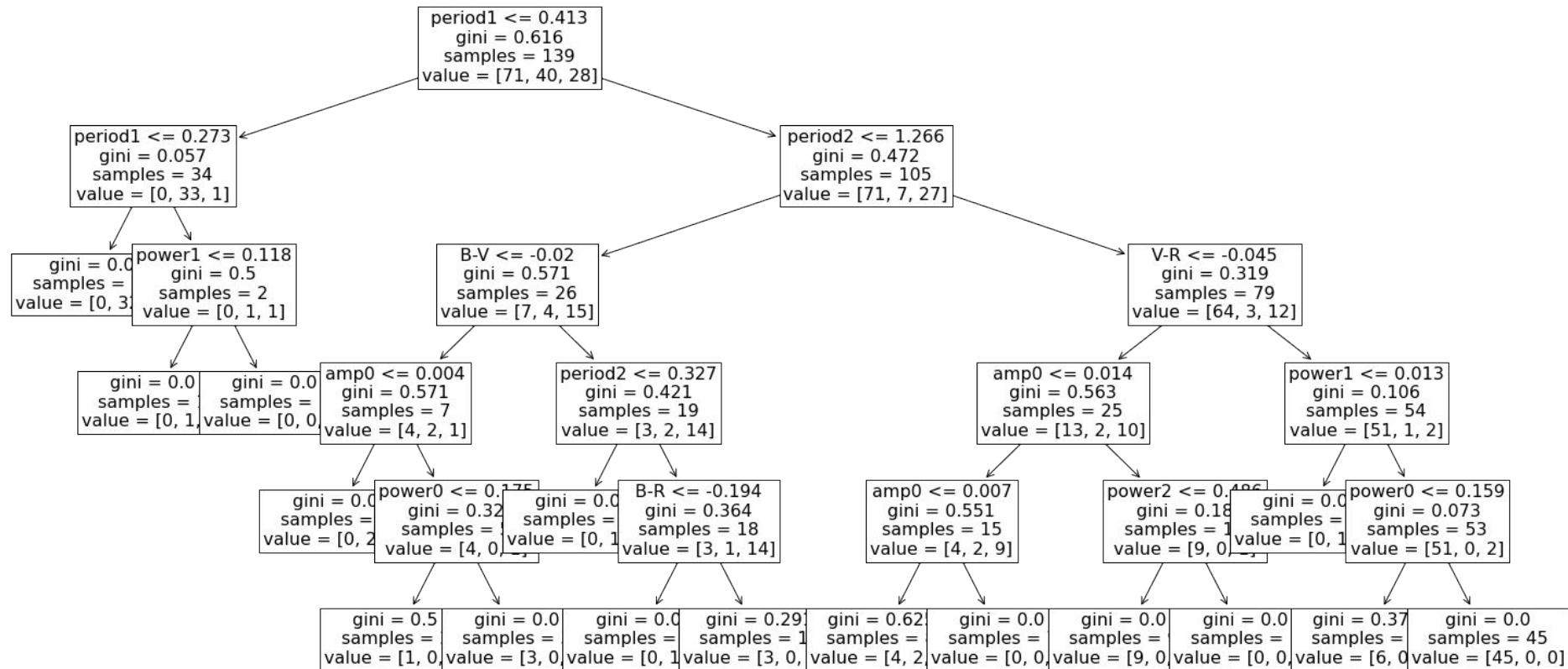
# Extra: Decision Tree with Different Levels 3



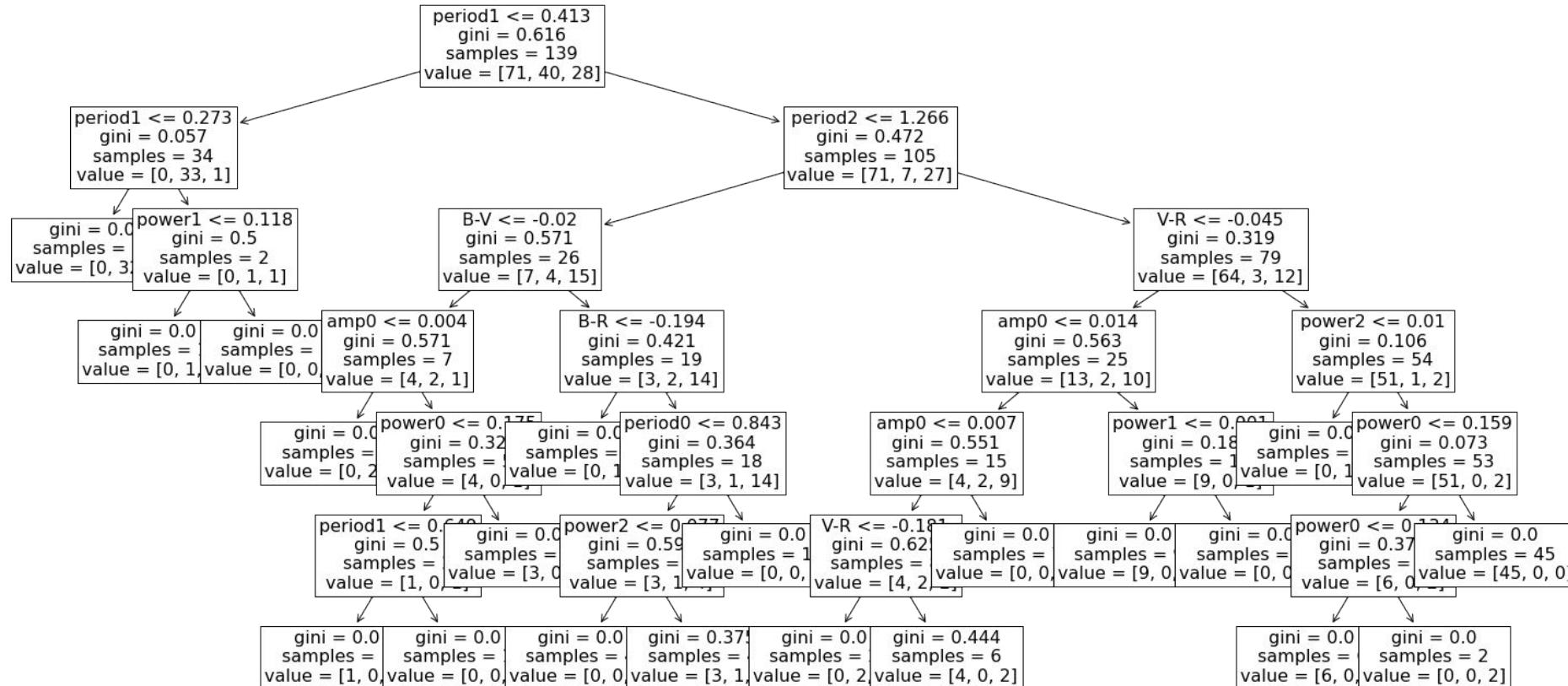
# Extra: Decision Tree with Different Levels 4



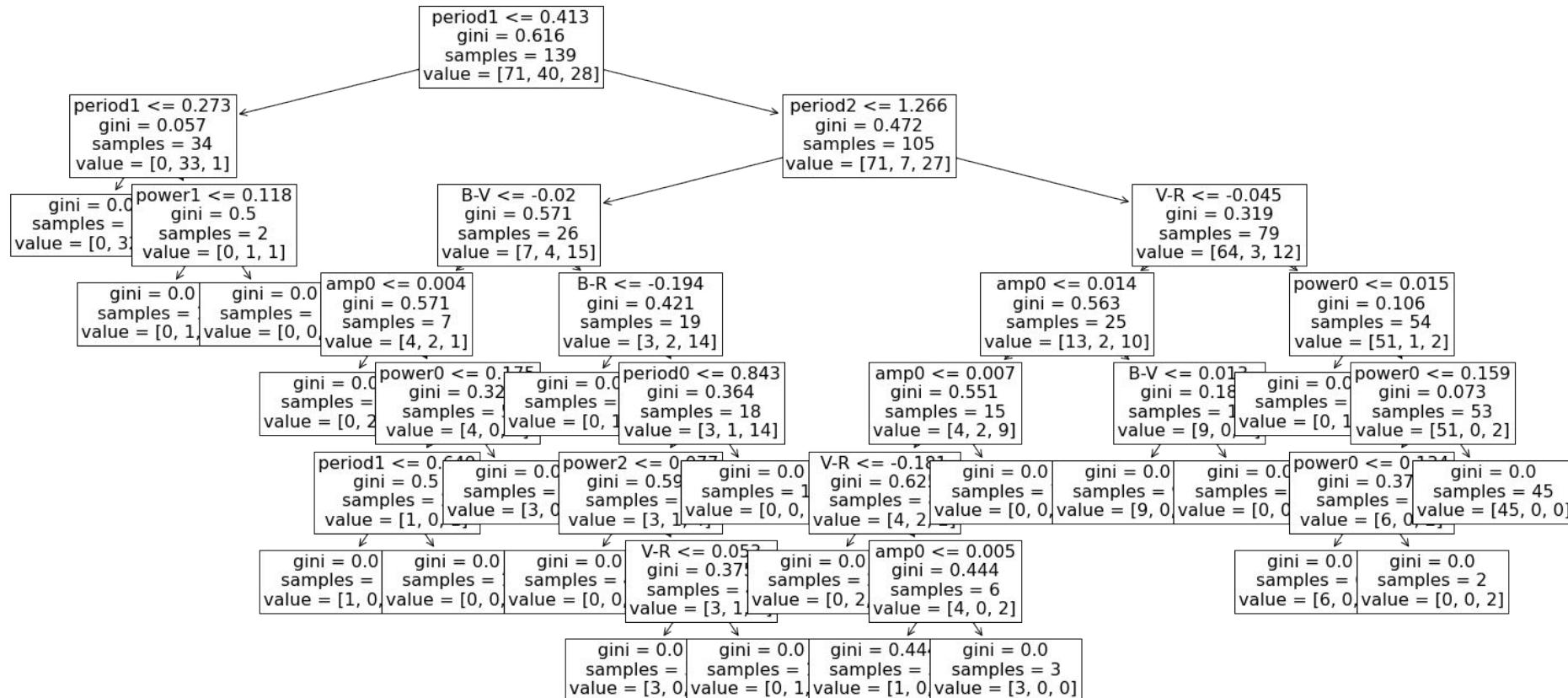
# Extra: Decision Tree with Different Levels 5



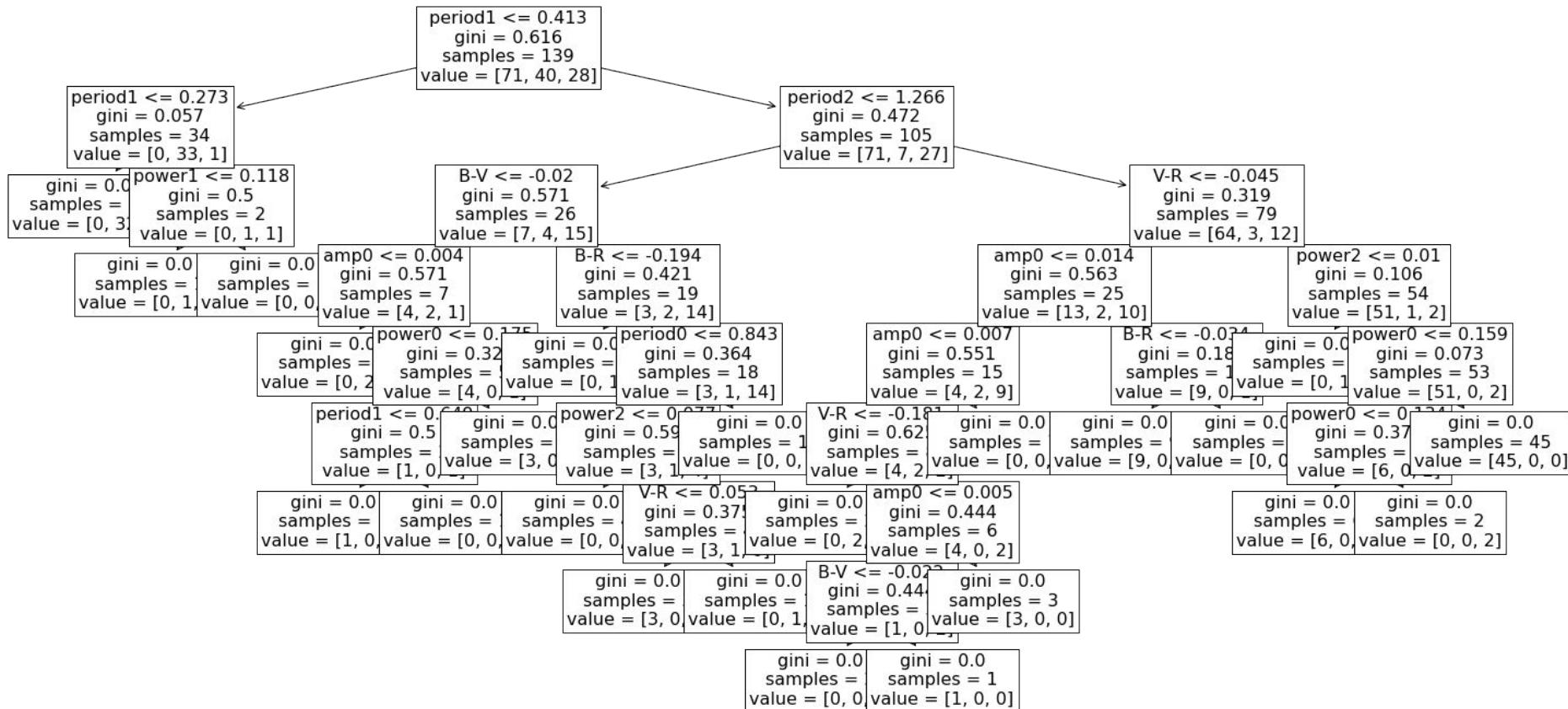
# Extra: Decision Tree with Different Levels 6



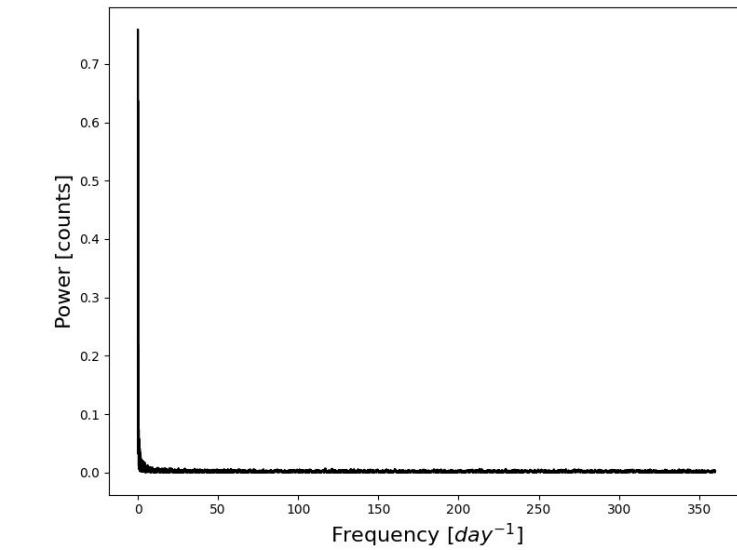
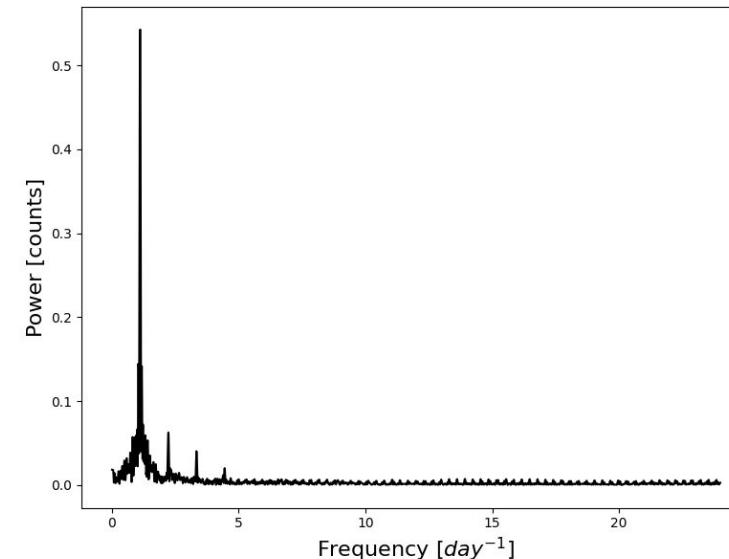
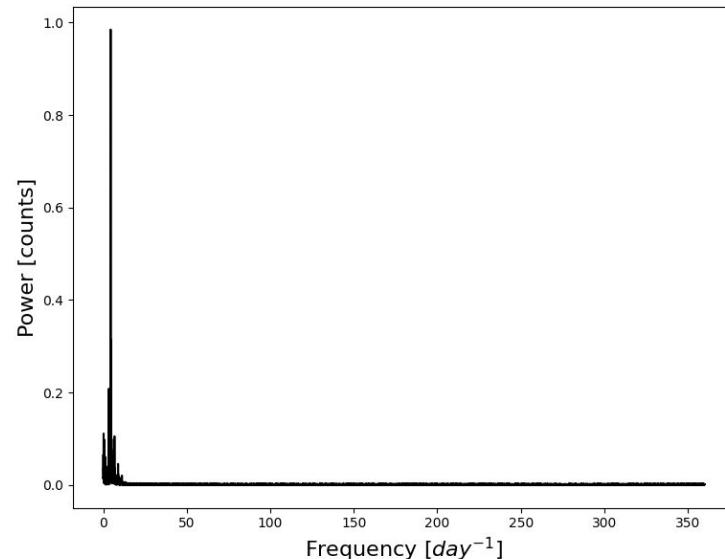
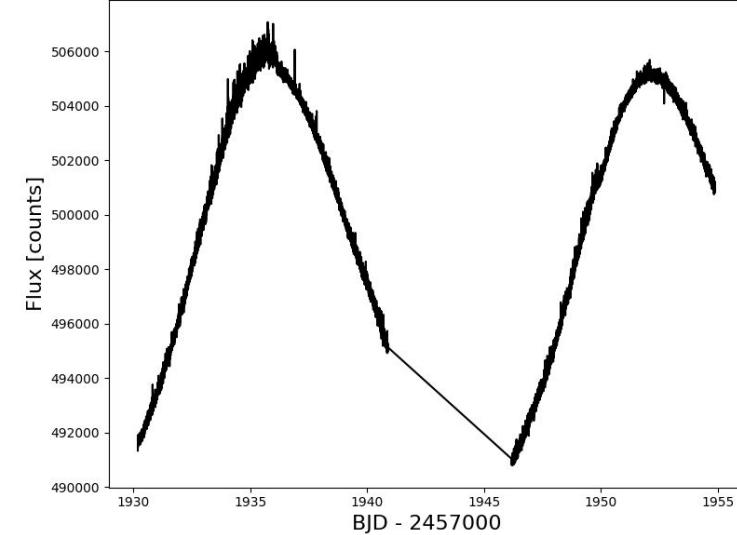
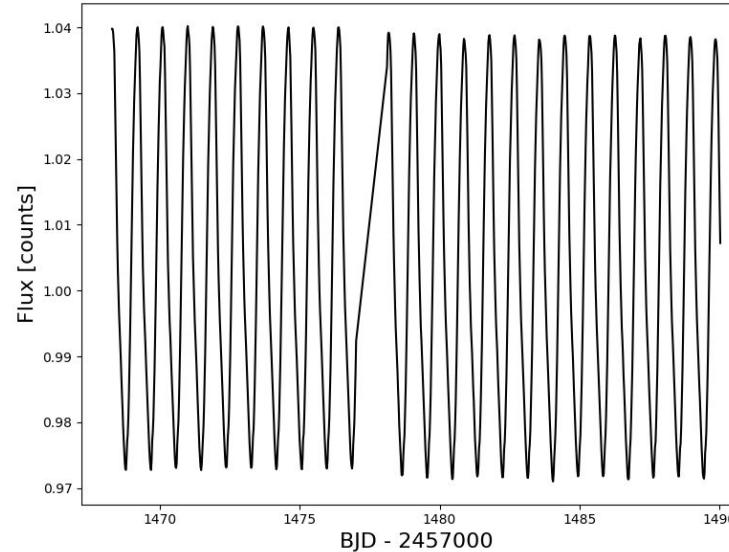
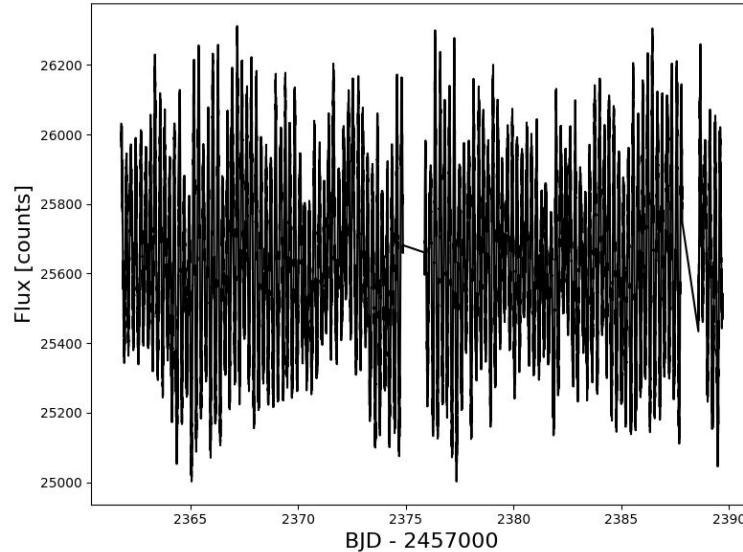
# Extra: Decision Tree with Different Levels 7



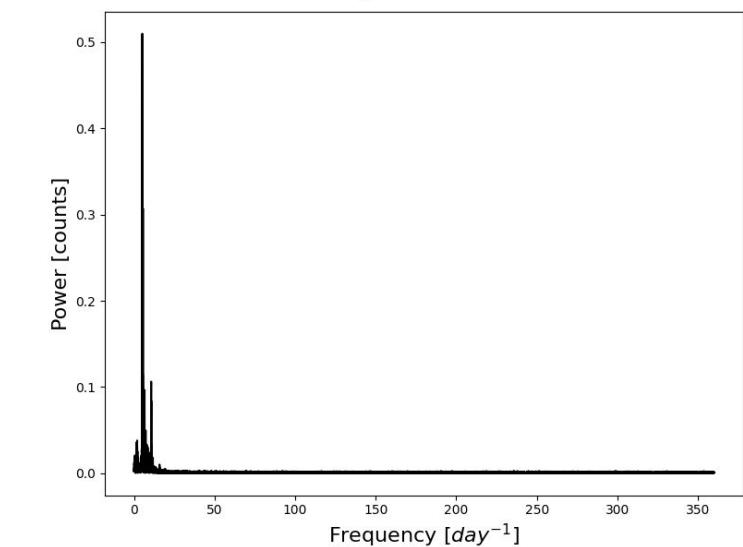
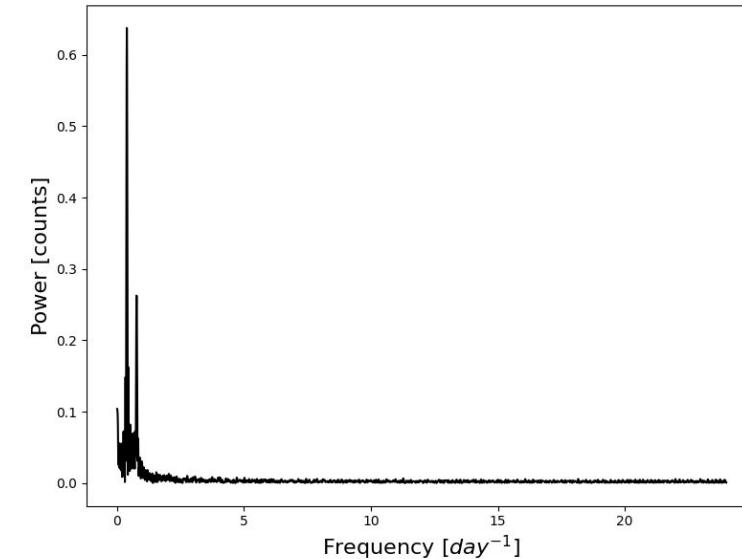
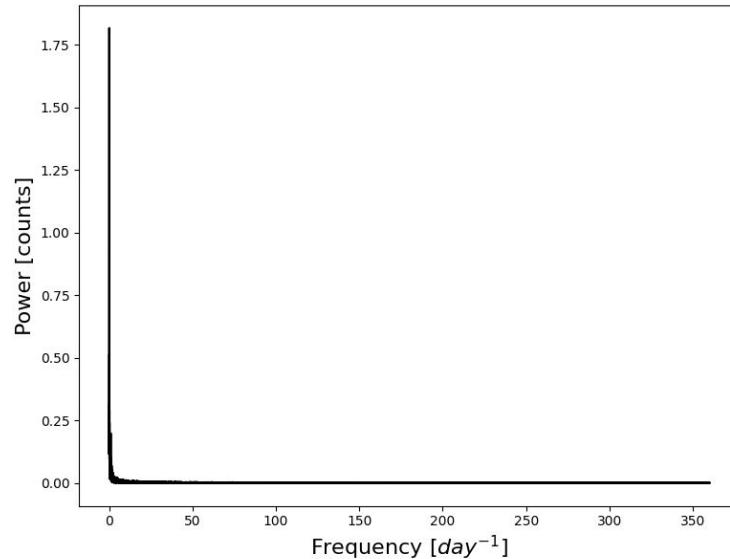
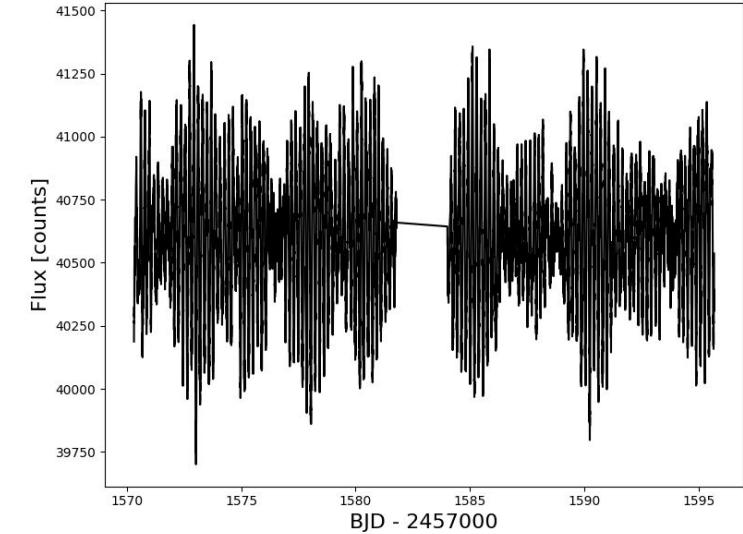
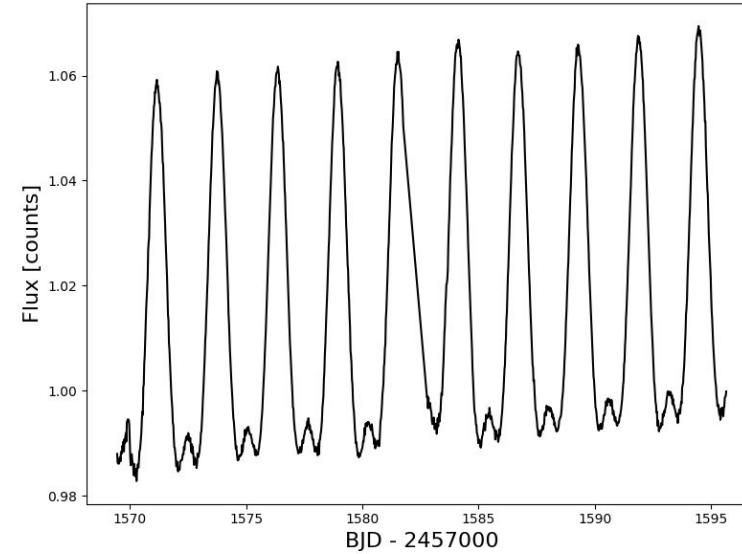
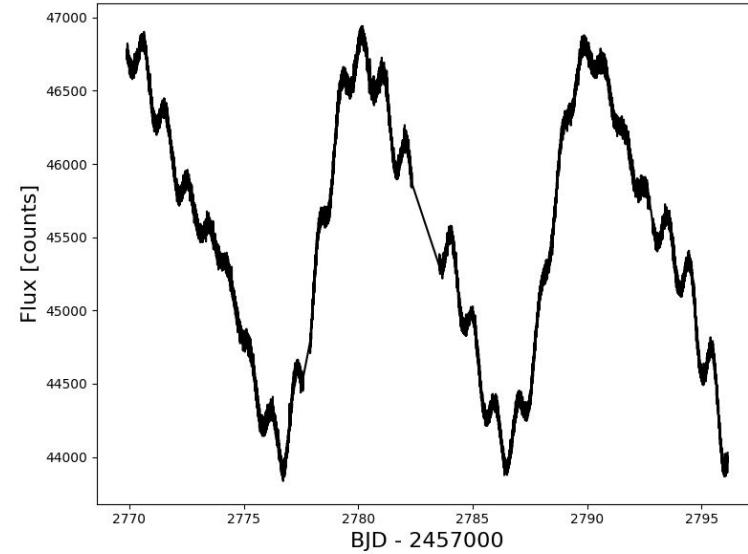
# Extra: Decision Tree with Different Levels 8



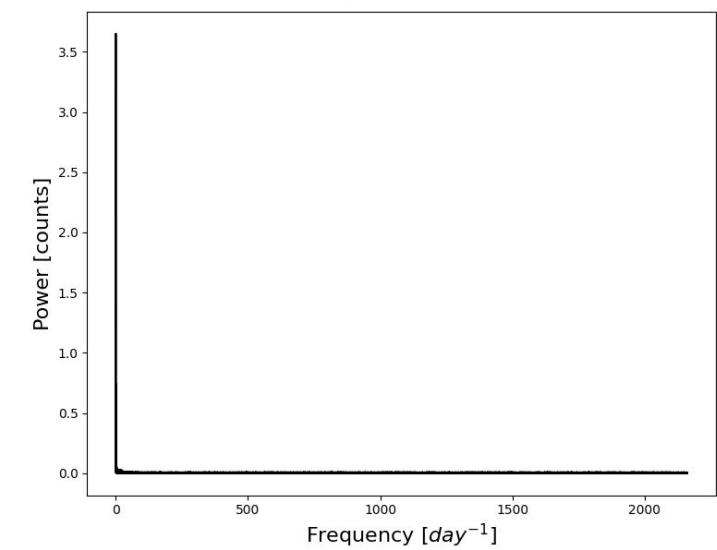
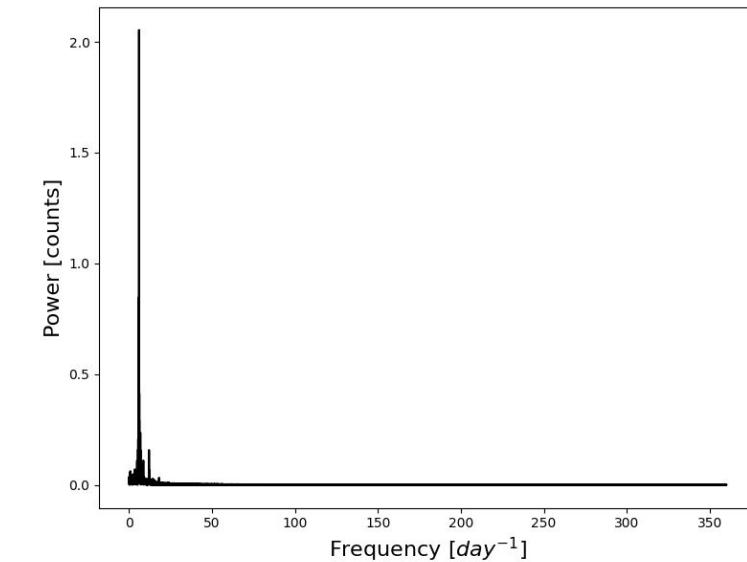
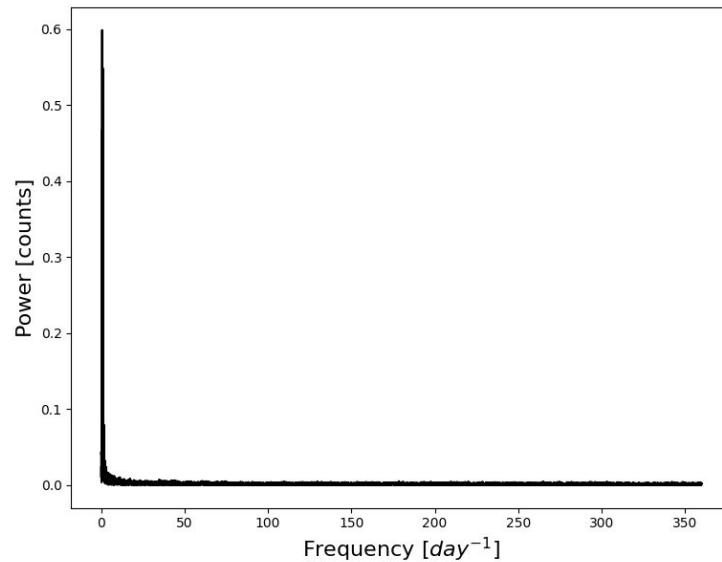
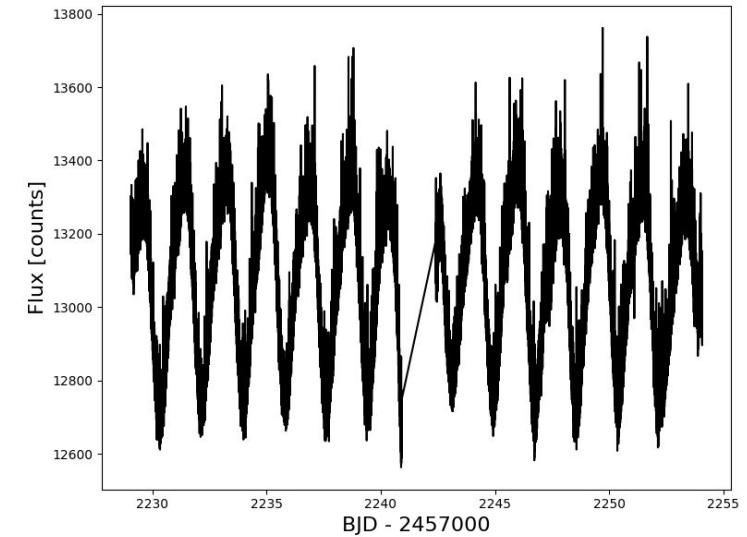
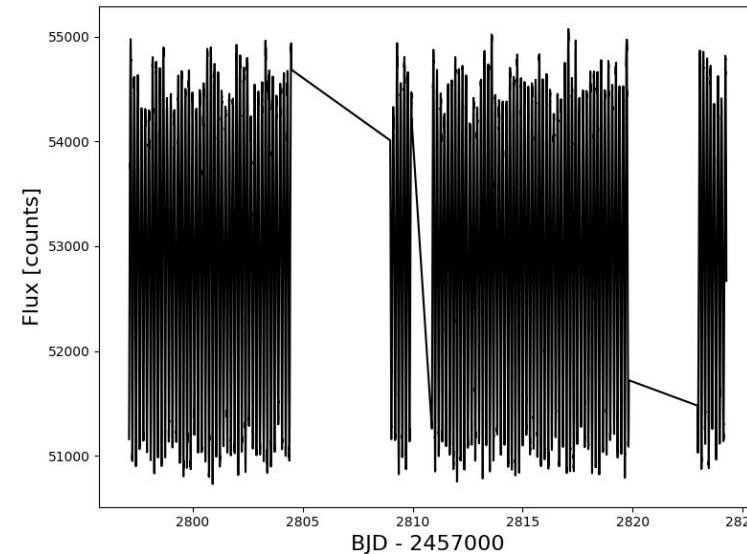
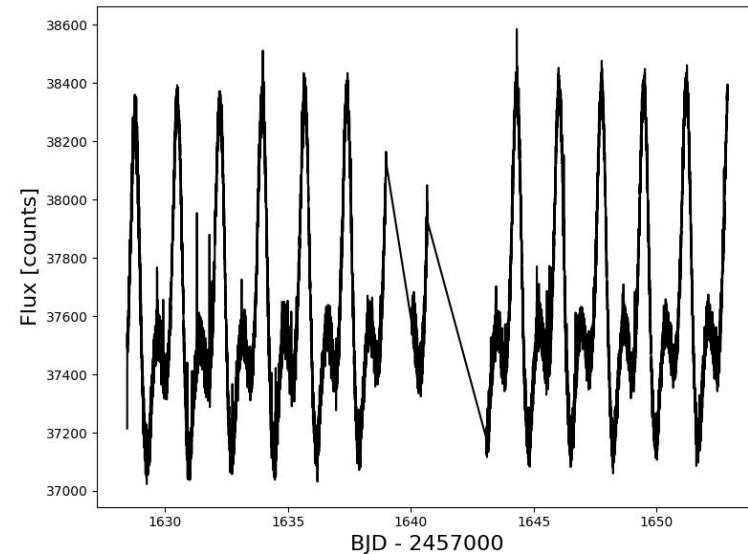
# Some Light Curves in the Sample



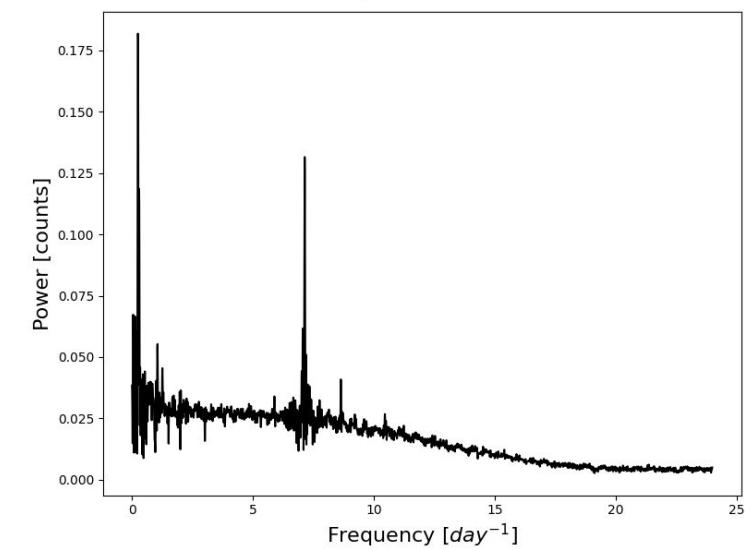
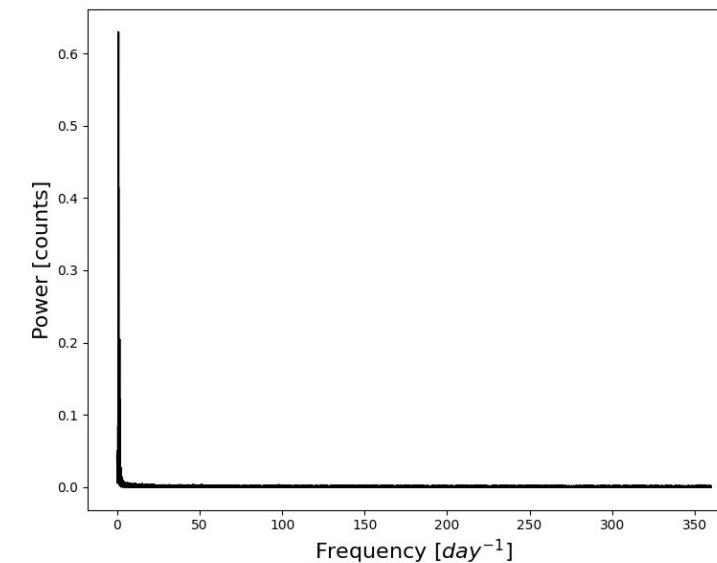
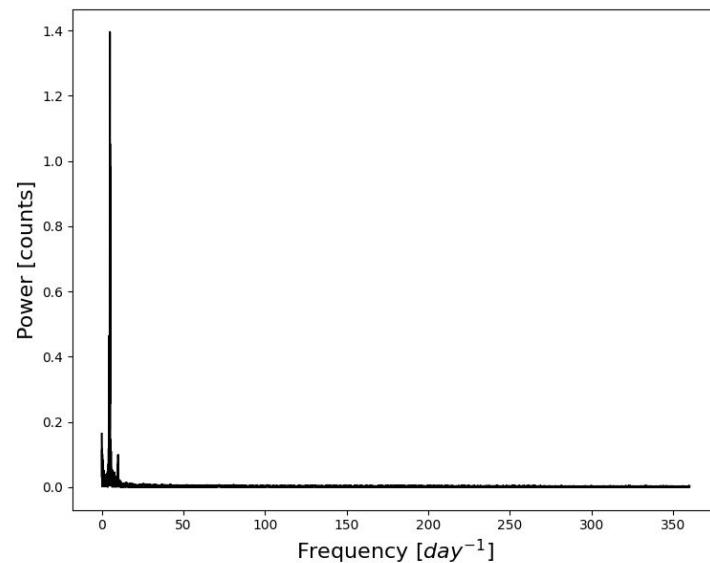
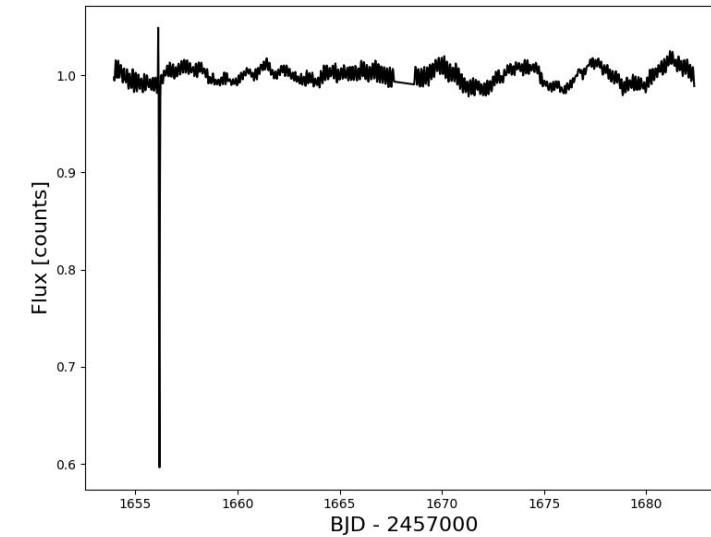
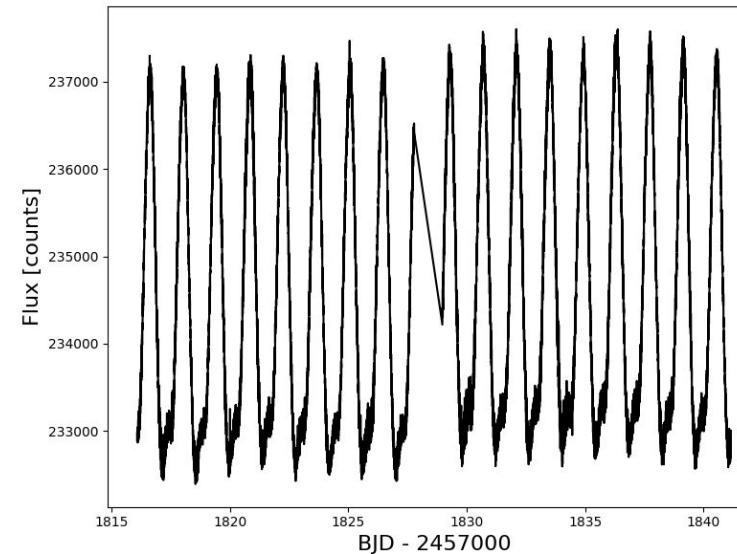
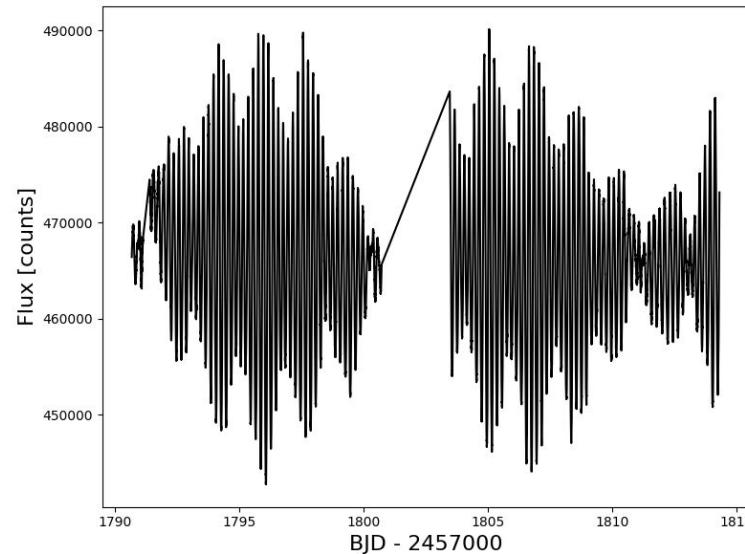
# Some Light Curves in the Sample



# Some Light Curves in the Sample



# Some Light Curves in the Sample



# Some Light Curves in the Sample

