CHIME Blazar Variability Analysis WWZ

2023-08-13

Evan Davies-Velie

Background

- Blazars are a subclass of active galactic nuclei (AGN) that emit highly variable radiation across the electromagnetic spectrum.
- Variability in blazars can occur on time scales ranging from minutes to years.
- CHIME gives us the unique opportunity to study short term variations in a large catalogue of blazars.
- Blazar variability is for the most part, stochastic, however, some quasi-periodic oscillations (QPOs) in blazars and other active galactic nuclei (AGNs) have been reported.

Motivation

- There is considerable uncertainty about the nature of the variable emission of blazars.
- By studying the variability of blazars, we can learn more about the structure and dynamics of AGN, as well as the physical mechanisms responsible for their emission.
- A blazar showing periodicity, may suggest that the blazar contains a binary supermassive black hole system and, in turn, would make it a candidate for gravitational wave research.
- The main goal of this project was constructing an efficient pipeline for detecting possible QPOs in blazar light curves.

Data Processing

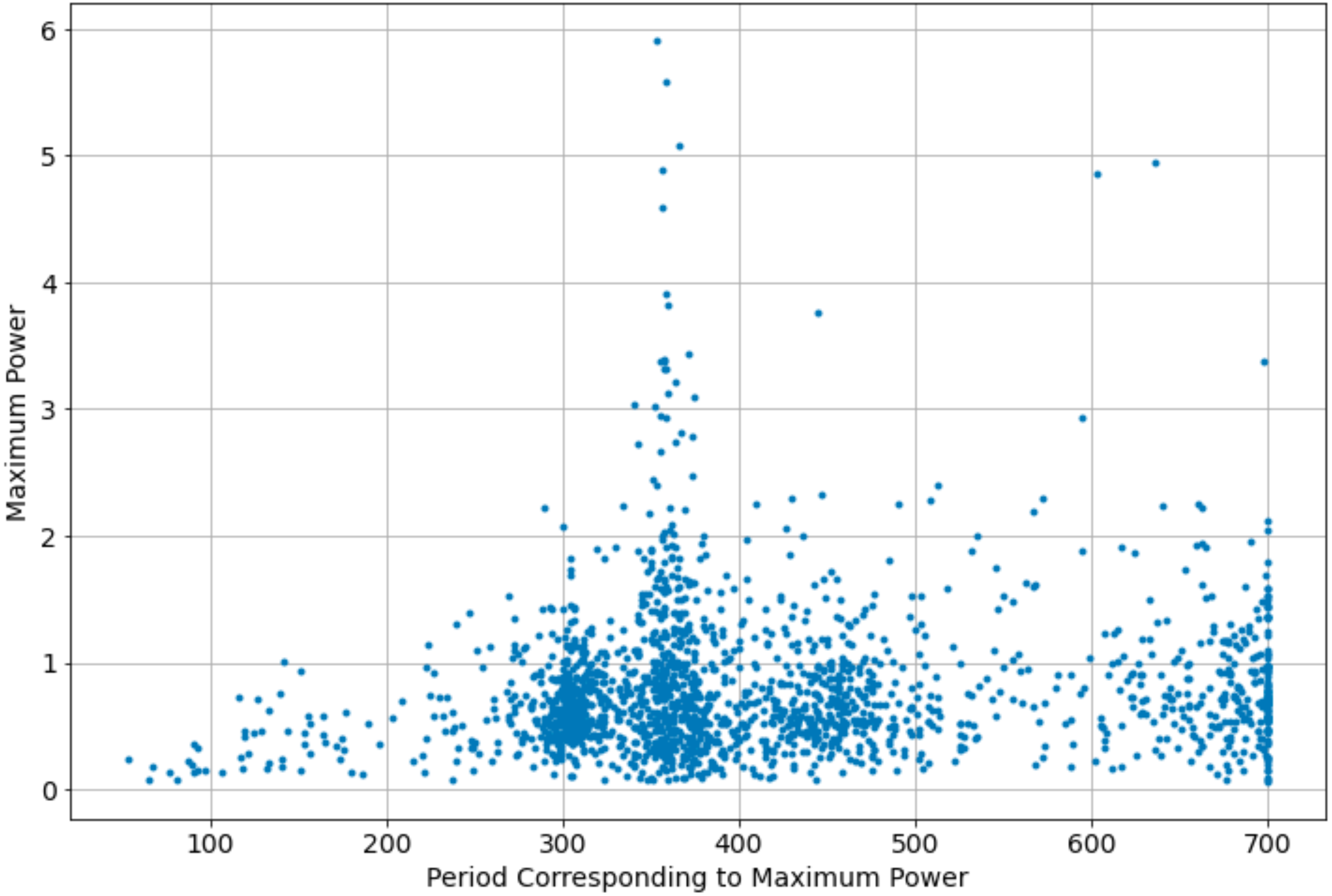
- I took the blazar daily flux data from revision 6, which contains 1847 blazars.
- Because we are focusing broad-band variations in total intensity, we averaged over frequencies and polarizations for each source. This gives us a single value of flux for each day.
- An inverse variance weighting is used for these averages.
- I also discarded daytime data, resulting in approximately 6-month gaps in the data.
- We normalize each light curve and also perform outlier removal using a sliding median filter.
- Because the data is no longer continuously sampled, standard Fourier techniques are not applicable.

Limit of 100 days

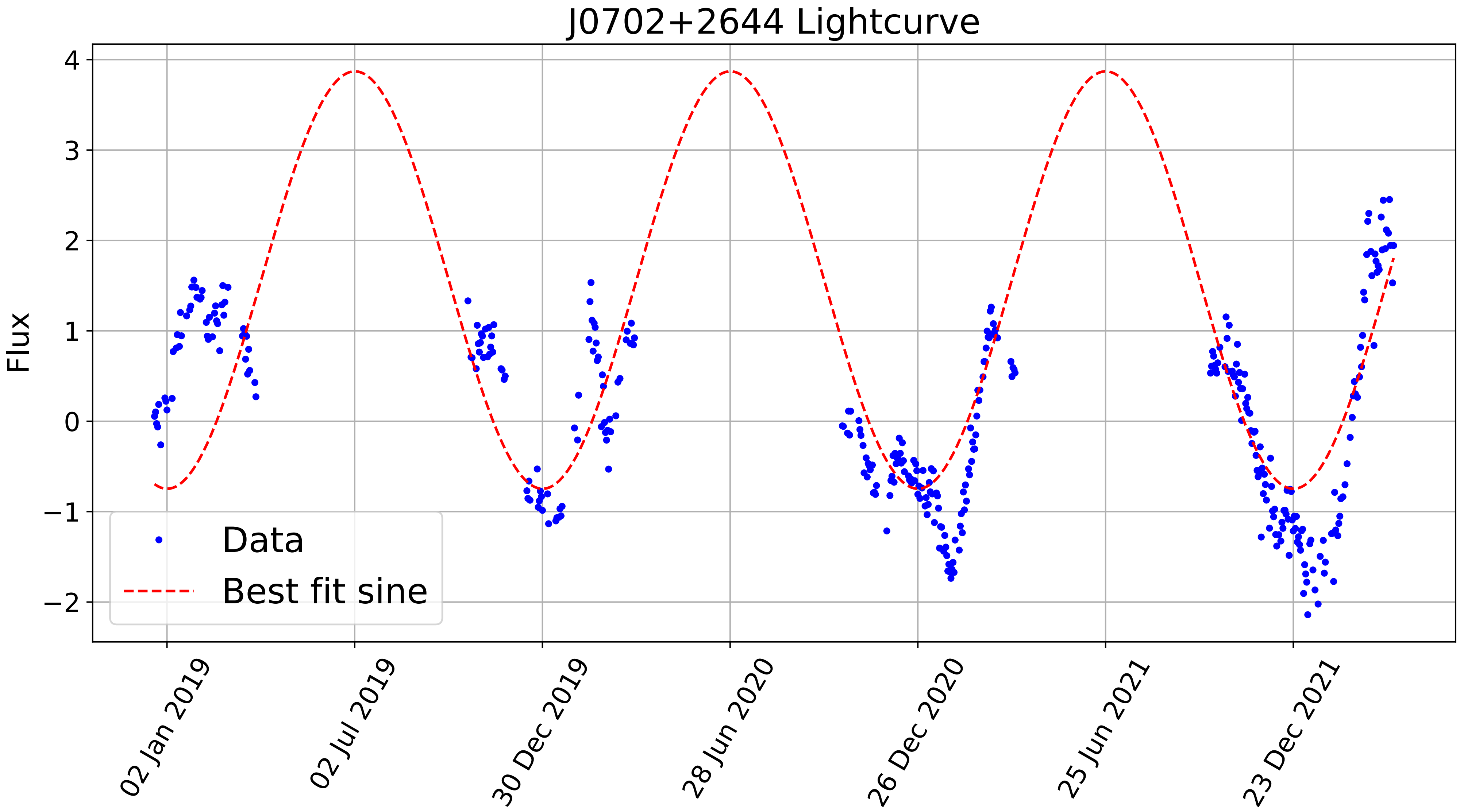
- During my research I came to realize that many sources exhibit very strong periodic variability with a timescale of almost exactly 1 year.
- This periodicity is not purely driven by the gaps in the data, but also the data itself.
- When looking at periods above 100 days, this effect caused many sources with periods of 365 or 180 days to be identified.
- By limiting the maximum period to 100 days, we hope to only find sources that truly exhibit periodic behavior.
- Short timescale variations is also where CHIME is advantageous over other telescopes, since we observe each source every single day.

Annual Periodicity from Lomb-Scargle Periodogram

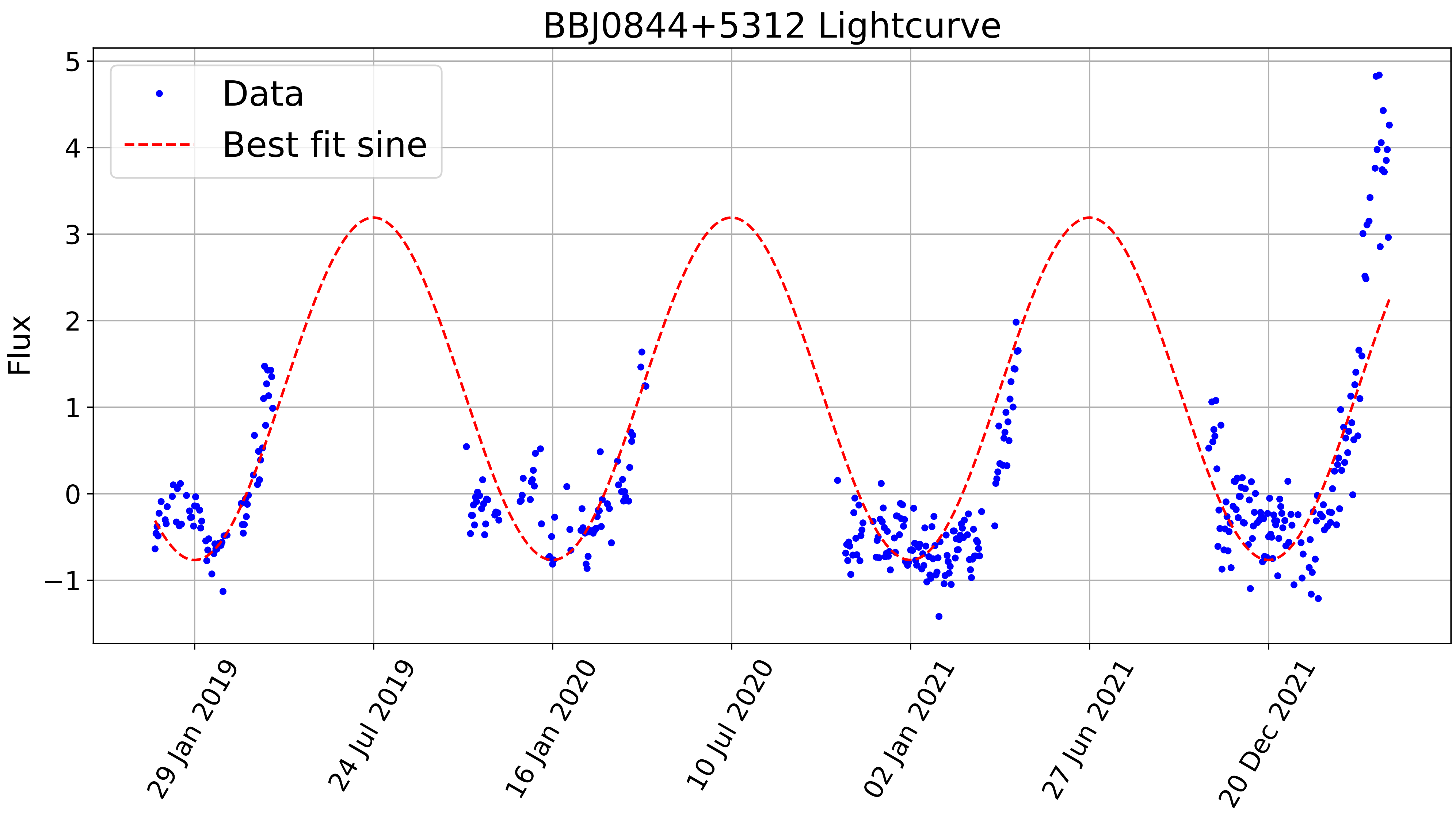
- This plot shows the maximum power achieved from the Lomb-Scargle periodogram and the period at which that peak occurs. Notice the dense, and tall group of points right around 365 days. On the next slides some of these source's lightcurves are shown.



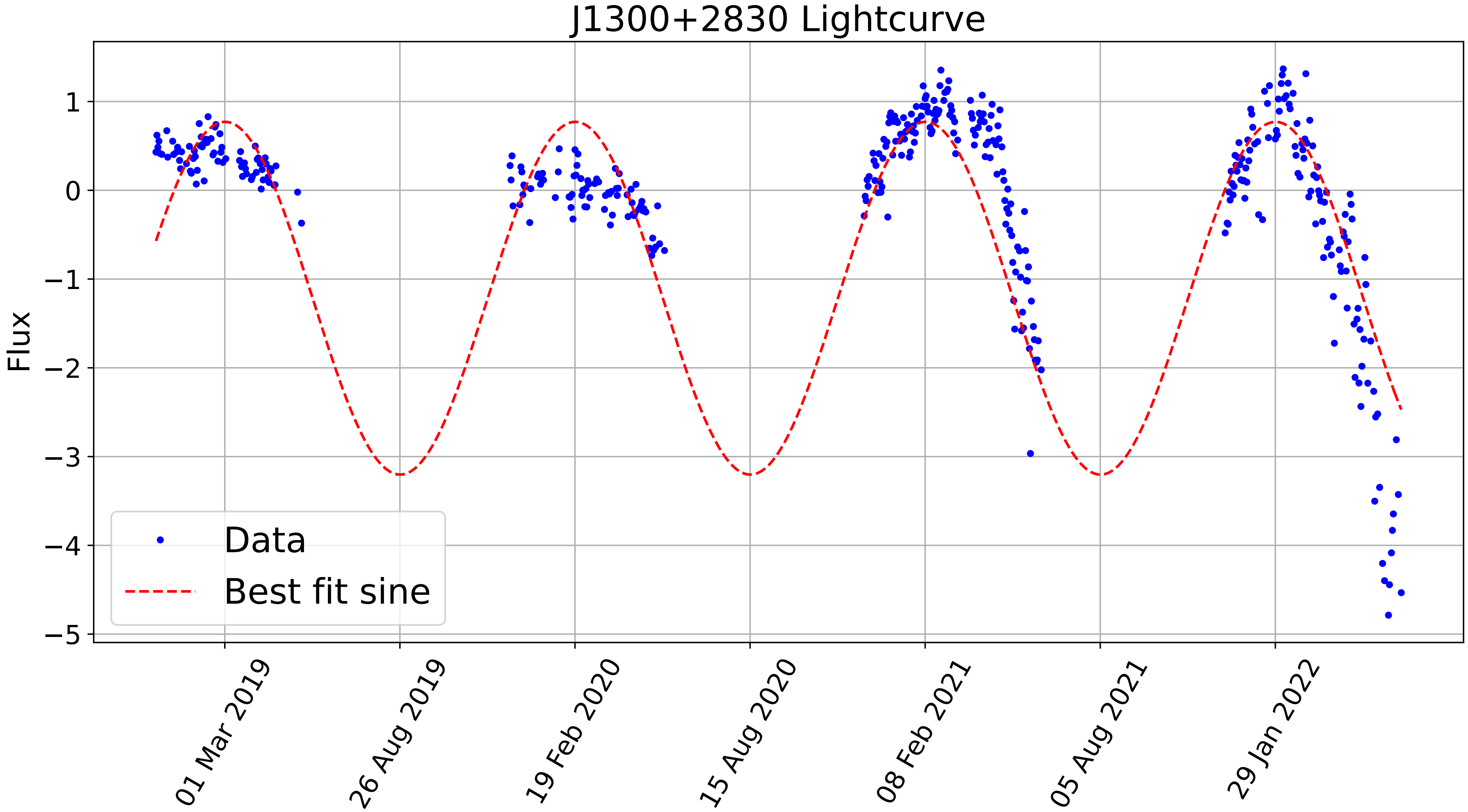
Example Sources with Annual Periodicity



Example Sources with Annual Periodicity



Example Sources with Annual Periodicity



Motivation for WWZ

- Wavelet transforms perform a similar function to the Fourier transform, but, in addition to this, they can break signals down into oscillations localized in space and time.
- However, when applied to unevenly sampled time series, the response of the wavelet transform is often more dependent on irregularities in the number and spacing of available data than on actual changes in the parameters of the signal.
- By casting the wavelet transform as a projection, we can derive its statistical behavior and devise advantageous rescaled transforms.
- By treating it as a weighted projection to form the weighted wavelet Z-transform (WWZ), we improve its ability to detect, and especially to quantify, periodic and pseudo-periodic signals.

Weighted Wavelet Z-Transform

- The Weighted Wavelet Z-Transform (WWZ) is a time-frequency analysis method, exploring both the frequency domain and the time domain.
- This has the advantage of allowing us to detect short-lived periodicity that may not be present for the entire duration of observation.
- One can think of it as being similar to a short-time Fourier transform using a Gaussian window function.
- The window size is frequency dependent and the parameter c controls how rapidly the window decays.
- c was set such that 3 periods can fit within the full width at half maximum (FWHM) of the Gaussian window.

Model Function

- We perform a weighted projection onto the three trial functions:

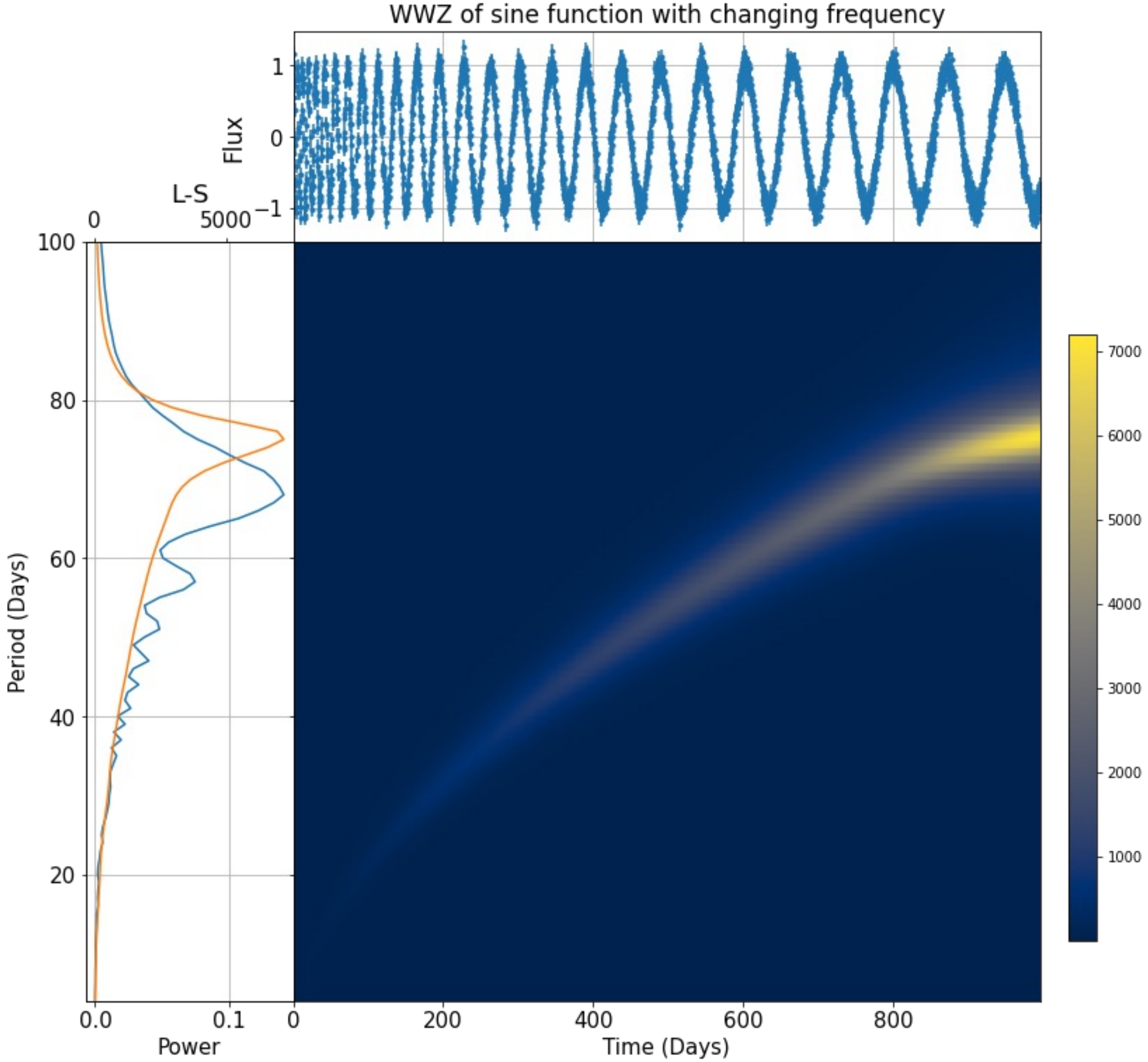
$$\phi_1(t) = \mathbf{1}(t),\ \phi_2(t) = \cos(\omega(t-\tau)),\ \text{and}\ \phi_3(t) = \sin(\omega(t-\tau))$$

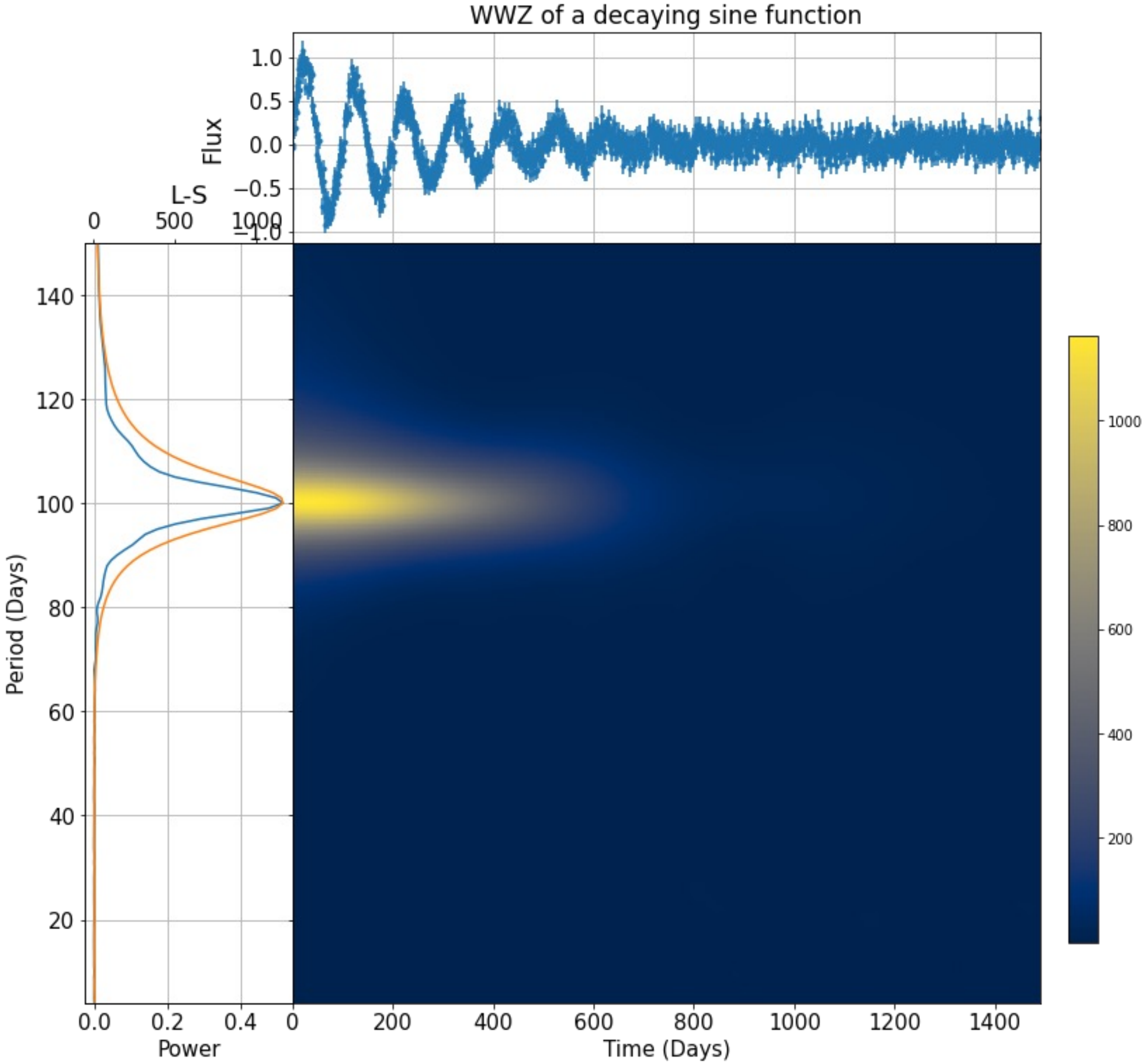
- Projection computes the coefficients y_a of the three trial functions for which the model function:

$$y(t) = \sum_a y_a \phi_a(t)$$

- best fits the data within the Gaussian window, in the sense that it minimizes the sum of the squared residuals.
- One modification we made to the method in the paper, is that we included the inverse variance weights of the data points from CHIME into the calculation.

Example WWZ with Mock Data





Estimating Peak Significance

- When analyzing blazar light curves, it is important to be cautious of false periodicities, which may arise due to random variability rather than a genuine periodic signal.
- We need a way to distinguish between true and false periodicities
- Generally, blazars show stochastic variability over a wide range of time scales. Therefore, in order to look for periodicity in the observed light curves, we want to compare them to a purely stochastic model.
- By simulating many stochastic lightcurves and comparing their WWZs with that of our observed lightcurves, we can determine how often a certain power level is exceeded just by chance and rule out blazars with peaks that don't surpass this.

Pipeline for WWZ Analysis

1

Generate a WWZ for the normalized observed blazar light curve.

2

Run an MCMC on the normalized light curve, constructing posterior distributions for σ , τ, and q.

3

Then draw from the posterior distributions of each parameter to simulate 100,000 light curves and generate a WWZ for each of them.

5

Take the maximum power achieved at each period over all simulations and tau values and compare it to the power achieved by the original WWZ at that period

6

If the power of the original WWZ exceeds the maximum power found in the previous step, label that period as significant. If at least two adjacent periods are significant, this source is a candidate for periodicity.

The Model: CAR(1) Process

- In this project, blazar light curves are modelled as a continuous time first-order autoregressive process (CAR(1)), which are commonly used to model AGN stochastic variability.
- For a CAR(1) process, the expected value of X(t) given X(s) for s < t is given by:

$$E(X(t)|X(s)) = e^{-\frac{\Delta t}{\tau}} X(s) + q\left(1 - e^{-\frac{\Delta t}{\tau}}\right)$$

and the variance in X(t) given X(s) is given by:

$$Var(X(t)|X(s)) = \frac{\tau\sigma^2}{2}\left(1 - e^{-\frac{2\Delta t}{\tau}}\right),$$

where Δt = t − s, σ is the variation amplitude at long timescale, and τ is the characteristic damping timescale, and q is the mean value of the light-curve

Estimating Model Parameters

- In order for a simulated lightcurve to represent our data we must estimate the parameters, σ, τ, and q.
- We set a uniform prior for the logarithm of the parameters σ and τ and a uniform prior for q.
- We then apply the Markov Chain Monte Carlo (MCMC) sampler emcee with 36 walkers and 5000 iterations, using the likelihood function shown on the next slide to generate posterior distributions for each of the parameters.
- We simulate full lightcurves, and then discard data such that the simulated lightcurve has the exact same sampling as the original lightcurve.

Likelihood Function For CAR(1) Process

We use the likelihood function used by others to model a CAR(1) process,

$$p(y_1, ..., y_n|\sigma, \tau, q) = \prod_{i=1}^{n}[2\pi(\Omega_i + \sigma^2)]^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\frac{(\hat{y}_i - y_i^*)^2}{\Omega_i + \sigma_i^2}\right],$$

where

$$y_i^* = y_i - q,$$

$$\hat{y}_i = e^{-(t_i - t_{i-1})/\tau}\hat{y}_{i-1} + \frac{e^{-(t_i - t_{i-1})/\tau}\Omega_{i-1}}{\Omega_{i-1} + \sigma_{i-1}^2}(y_{i-1}^* - \hat{y}_{i-1}),$$

and

$$\Omega_i = \Omega_1(1 - e^{-2(t_i - t_{i-1})/\tau}) + e^{-2(t_i - t_{i-1})/\tau}\Omega_{i-1}\left(1 - \frac{\Omega_{i-1}}{\Omega_{i-1} + \sigma_{i-1}^2}\right).$$

The initial values are $\Omega_1 = \frac{\tau\sigma^2}{2}$ and $\hat{x}_1 = 0.$ We set a uniform prior for the logarithm of the parameters σ and τ and a uniform prior for q.

Significance Criteria

- For each source, we simulate 100,000 lightcurves, with parameters drawn from the posterior distributions gathered from the MCMC.
- For each period, we take the maximum power across all simulated lightcurves across all time/tau pixels.
- If the power achieved at that period in our original lightcurve surpasses this, we label the source as significant.
- Since we are looking at 100 periods for each source, and there are 1847 sources we have:

$$1847 \cdot 100 \cdot \frac{1}{100000} \approx 1.85$$

- Meaning we should expect around 2 false alarms. If more than 2 individual periods are identified, then they are significant at face value using Gaussian statistics and should therefore be studied further.

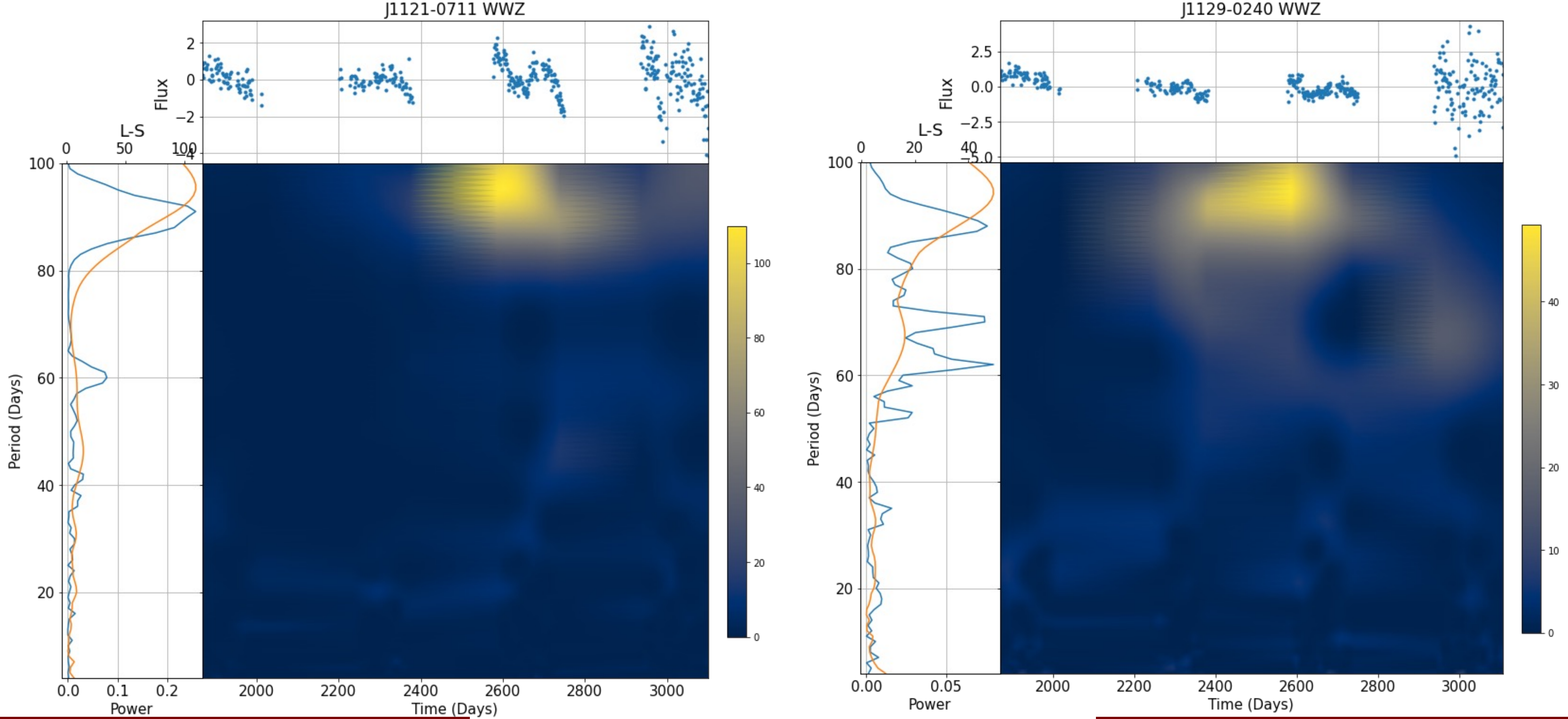
Significance Criteria

- For each source, we simulate 100,000 lightcurves, with parameters drawn from the posterior distributions gathered from the MCMC.
- For each period, we take the maximum power across all simulated lightcurves across all time/tau bins.
- If the power achieved at that period in our original lightcurve surpasses this, we label the source as significant.
- Assuming the light curve is properly represented by our simulations and that noise at different periods is uncorrelated, then we expect there to be a 1/100,000 chance that the observed power will exceed all simulations in each period bin.
- There are 100 period bins, therefore there is a 1/1000 chance that a particular light curve will be a false alarm.
- Because there 1847 sources, then we expect roughly 2 false alarms.

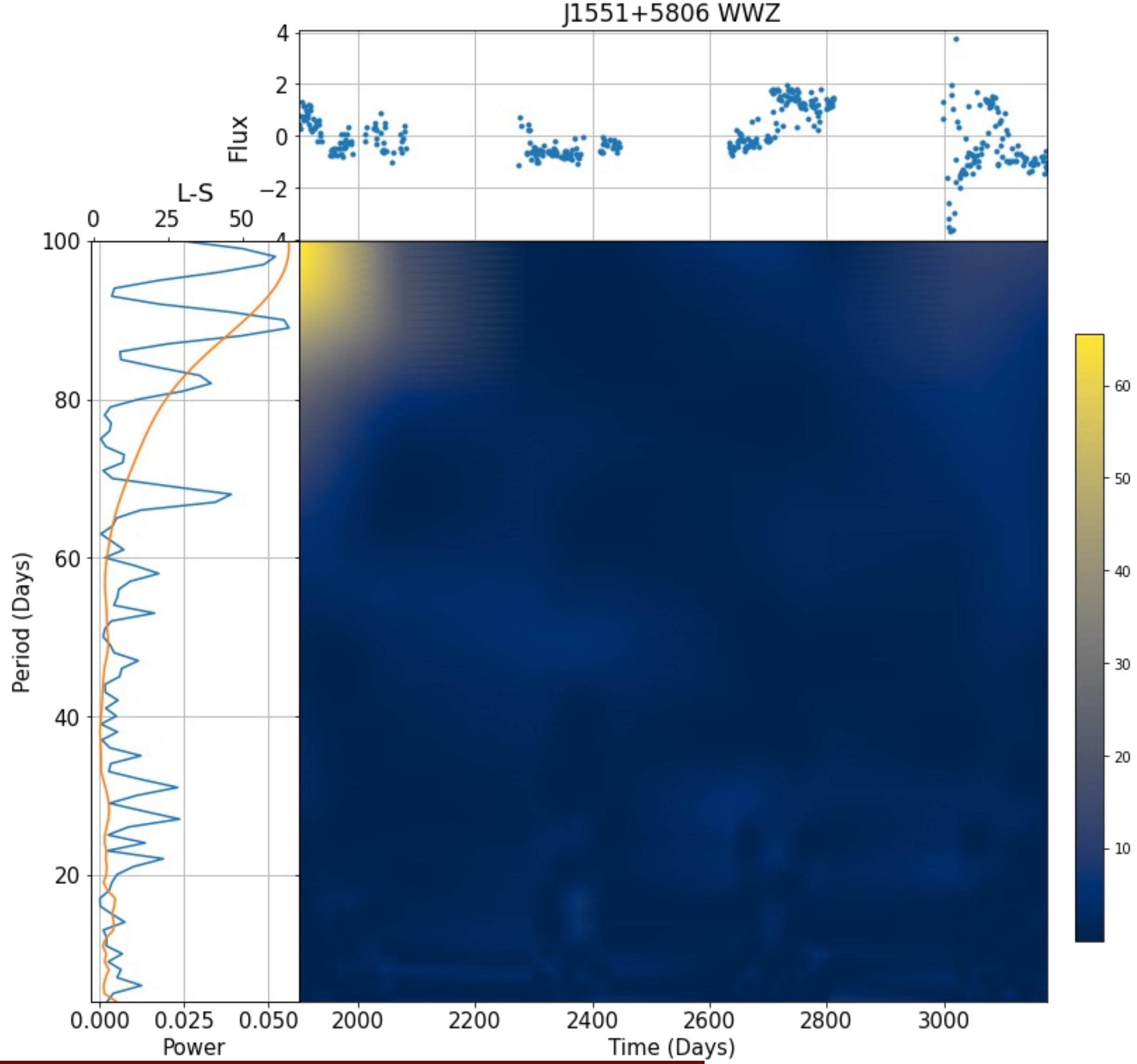
Results

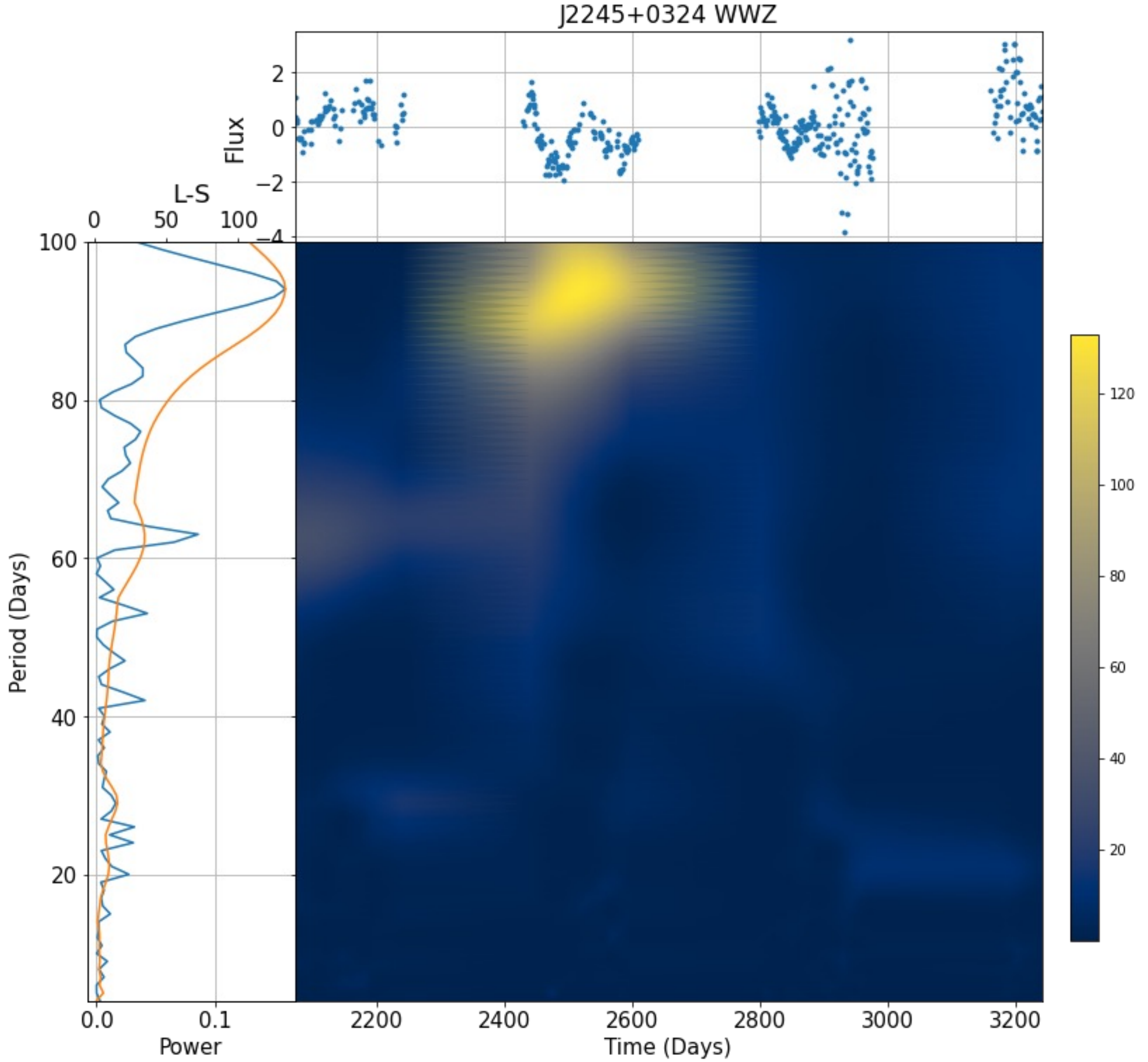
- Running the pipeline on all 1847 sources identifies 6 sources with significant periodicity under 100 days.
- Despite the width of our Gaussian window being 3 periods wide, when inspected by eye, these sources do not seem to contain 3 full oscillations.
- Furthermore, all of these sources are significant at a timescale of around 90 days, which is a harmonic of 1 year.

Significant Sources

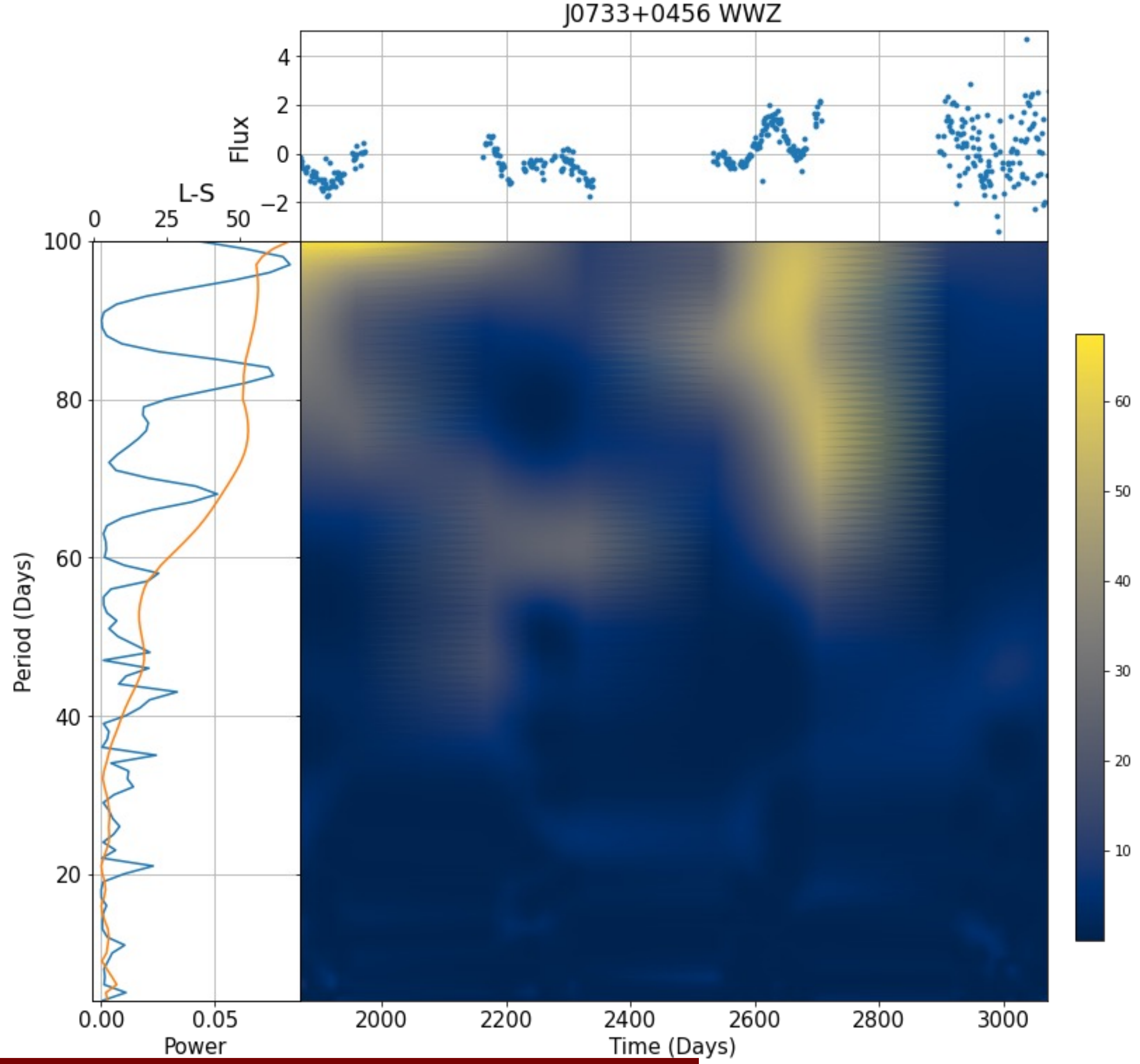


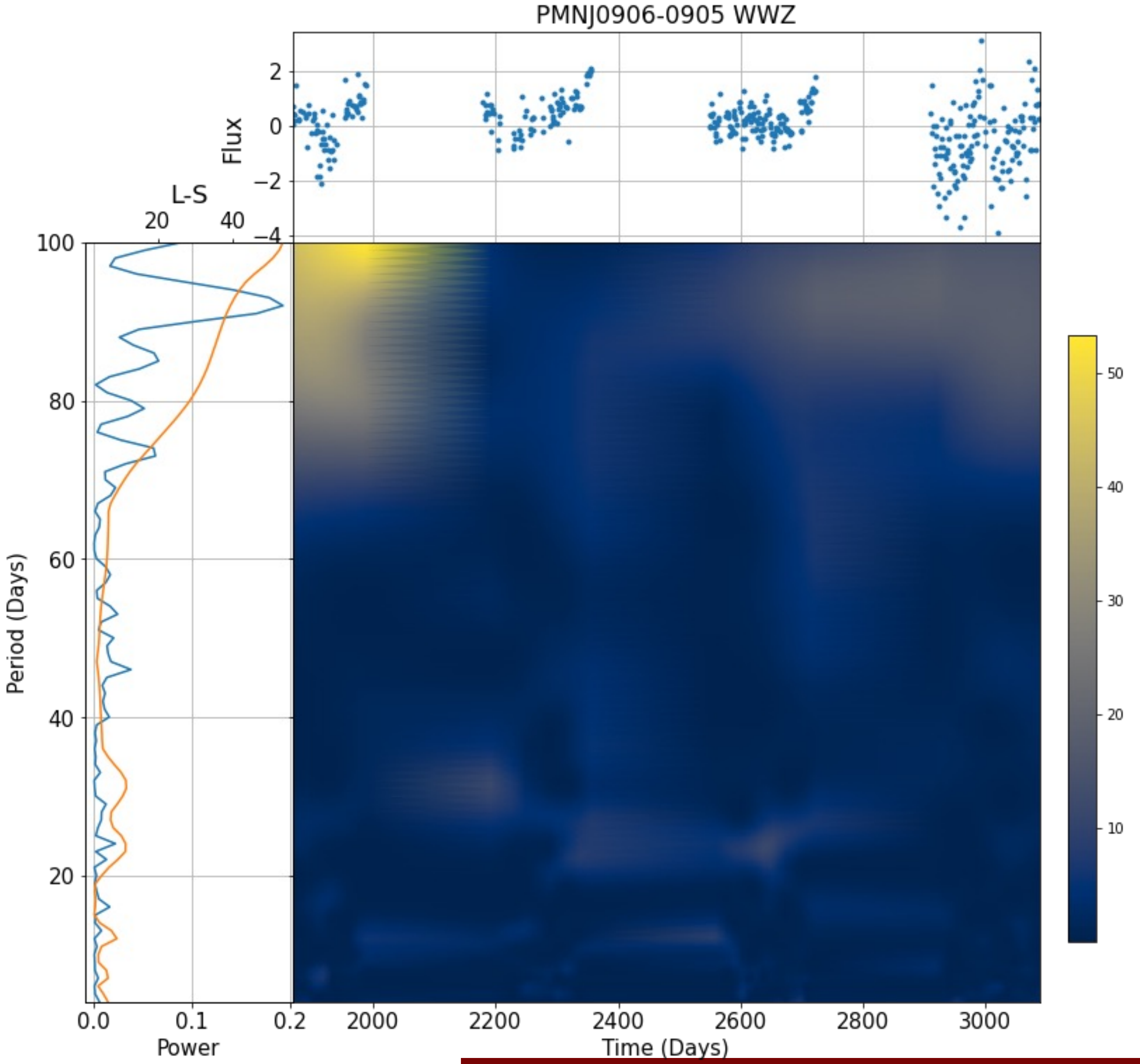
Significant Sources





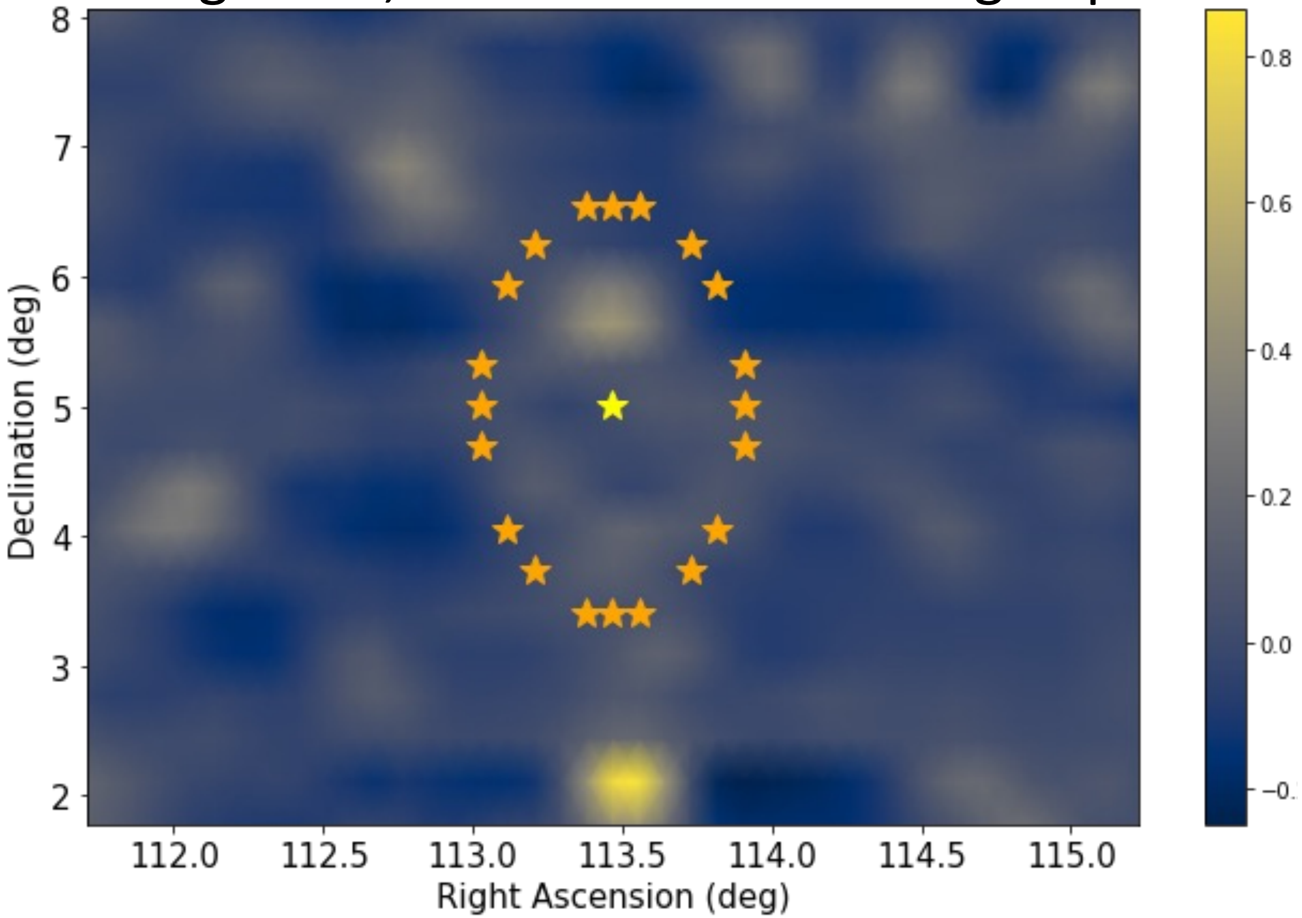
Significant Sources

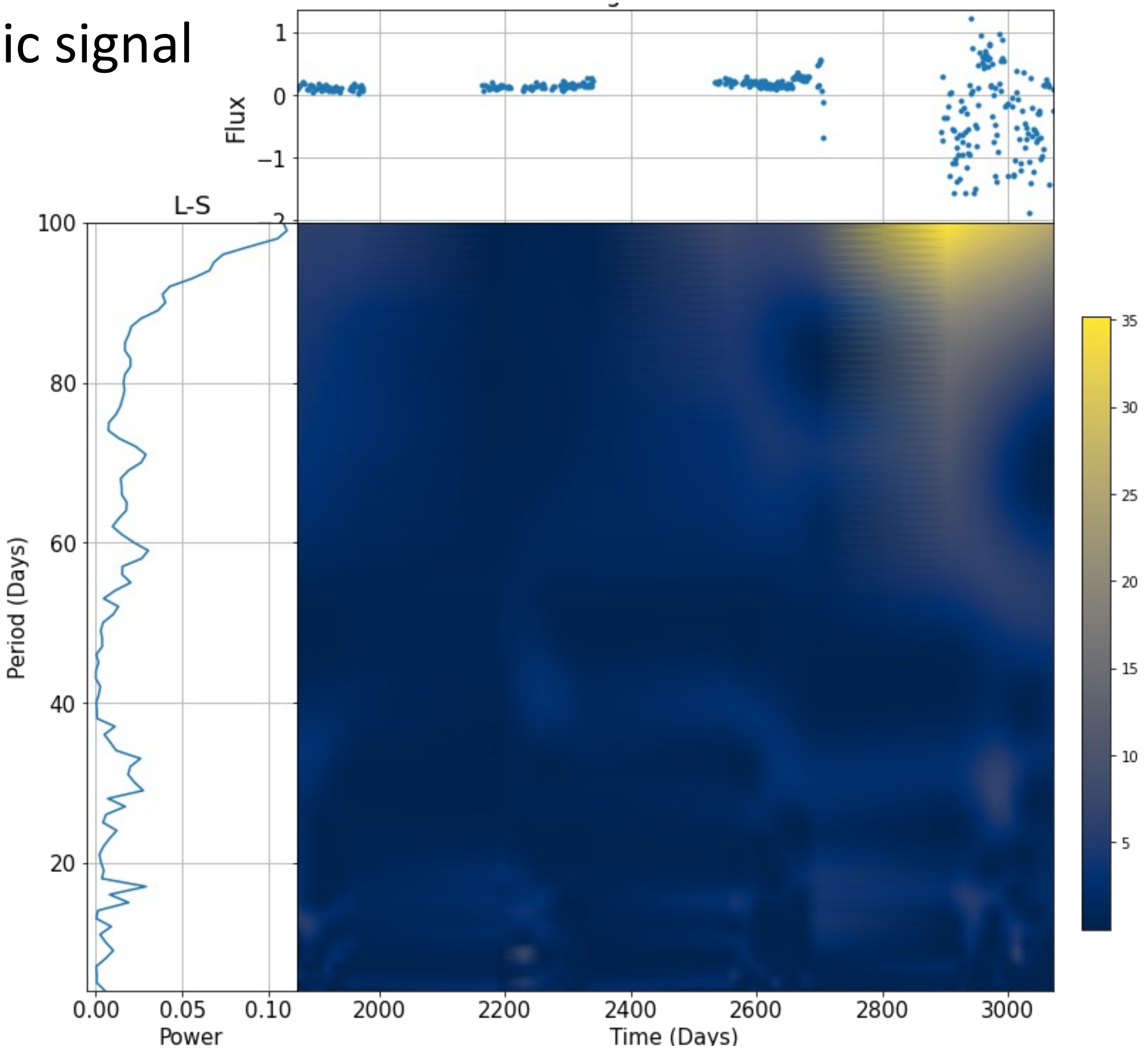




Ringmap Analysis Example (J0733+0456)

- In order to determine whether or not the periodic signal is coming from the source or something in the background, we can look at the ringmap.





Conclusion

- Even though the ringmap analysis showed that the significant periodic signal detected by the pipeline was characteristic of the source and not the background it is still not convincing that these sources truly exhibit QPOs.
- All the sources are consistent with stochastic variation, with a few exceptions that seem to be affected by annual (365 day) variation.
- Understanding the source of this variation will be important if we ever want to extend this analysis to longer periods.
- More data is necessary to look for longer period QPOs, as we only barely have enough data to look for QPOs with periods over 1 year with our 3-period requirement.