

KRITTIKA SUMMER PROJECTS 2024

# Searching for Periodicity in AGN light curves

Fida Fathima T P, Haemanth Ruban, Samarth Majumdar, Sameer Patil,  
Avinash Kumar Paladi, and Kenil Ajudiya



KRITTIKA SUMMER PROJECTS 2024

# Searching for Periodicity in AGN light curves

Fida Fathima T P<sup>1</sup>, Haemanth Ruban<sup>2</sup>, Samarth Majumdar<sup>3</sup>, Sameer Patil<sup>4</sup>,  
Avinash Kumar Paladi<sup>5,6</sup>, and Kenil Ajudiya<sup>5,6</sup>

<sup>1</sup>Jawaharlal Nehru University, New Delhi

<sup>2</sup>IIT Hyderabad

<sup>3</sup>University of Hyderabad

<sup>4</sup>IIT BHU, Varanasi

<sup>5</sup>Indian Institute of Science, Bangalore

<sup>6</sup>Mentors for this project

Copyright © 2024 Krittika IITB  
PUBLISHED BY KRITTIKA: THE ASTRONOMY CLUB OF IIT BOMBAY  
[GITHUB.COM/KRITTIKAIITB](https://github.com/KrittikaIITB)  
First Release, August 2024

# Abstract

Supermassive black holes (SMBHs) are the largest type of black hole with masses ranging from thousands to billions of times the mass of the sun ( $M_{\odot}$ ). AGN are compact regions at the centres of galaxies powered by SMBHs which emit large amounts of electromagnetic radiation often outshining the light of billions of stars. Quasars are a subclass of AGN, being some of the brightest objects in the universe, and are some of the farthest objects from the earth, with the nearest being millions of lightyears away.

Supermassive black hole binaries (SMBHBs) are systems of two SMBHs at centres of galaxies orbiting each other. They are said to be products of galaxy mergers, and emit gravitational waves of very low frequencies which prevent current gravitational wave detectors from detecting them.

In this study, we focus on simulating and studying the light curves of AGN as observed by time-domain photometric surveys. Specifically, we try to find periodicities in light curve data, which may indicate the presence of SMBHBs. We simulate light curves of current and future surveys such as the LSST and CRTS, over large characteristic timescales ranging over several years using a Damped Random Walk (DRW) model.

Using a mix of simulated and real data, we test out four methods of finding periodicities in the light curves - the Generalised Lomb-Scargle Periodogram (GLSP), Weighted Wavelet Z - Transform (WWZ Transform), REDFIT, and Bayesian Inference. Each of these methods have their advantages and disadvantages as we will see later in the study.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	<b>Super Massive Black Holes</b>	5
1.2	<b>Active Galactic Nuclei</b>	5
1.3	<b>Super Massive Black Hole Binaries</b>	6
1.3.1	Inspiral	6
1.3.2	Merger	6
1.3.3	Ringdown	6
1.4	<b>Observational Techniques</b>	7
1.4.1	Gravitational Wave Observatories	7
1.4.2	<b>Electromagnetic Observations</b>	7
1.5	<b>Objectives</b>	7
<b>2</b>	<b>Methods</b>	<b>9</b>
2.1	<b>Simulating Data</b>	9
2.1.1	Damped Random Walk	9
2.1.2	Injecting a sinusoid	10
2.2	<b>Analysis</b>	10
2.2.1	Generalized Lomb-Scargle Periodogram	11
2.2.2	Weighted wavelet Z-transform	11
2.2.3	REDFIT	11
2.2.4	Bayesian Inference	12
<b>3</b>	<b>Results</b>	<b>13</b>
3.1	<b>Observational Data</b>	13
3.2	<b>Lomb Scargle Periodogram</b>	14

<b>3.3</b>	<b>WWZ Results</b>	<b>17</b>
<b>3.4</b>	<b>REDFIT Results</b>	<b>17</b>
<b>3.5</b>	<b>Bayesian Inference Results</b>	<b>19</b>
3.5.1	Bayes Analysis of PG1302 CRTS Observations .....	19
3.5.2	Bayes Analysis of Simulated LSST light curve data for a SMBHB .....	20
<b>4</b>	<b>Conclusion .....</b>	<b>23</b>
<b>4.1</b>	<b>Comparison of Results</b>	<b>23</b>
4.1.1	Accuracy .....	23
4.1.2	Time taken for computation .....	23
4.1.3	Applicability .....	23
<b>4.2</b>	<b>Summary</b>	<b>24</b>
<b>5</b>	<b>Bibliography .....</b>	<b>25</b>
	<b>Bibliography .....</b>	<b>25</b>
	<b>Books</b>	<b>25</b>
	<b>Articles</b>	<b>25</b>
	<b>Inproceedings</b>	<b>26</b>
	<b>Software</b>	<b>26</b>
	<b>Misc</b>	<b>26</b>



# 1. Introduction

## 1.1 Super Massive Black Holes

Super Massive Black Hole (SMBH) is the largest type of black hole, with the mass of more than hundred thousand mass of the sun ( $M_{\odot}$ ). Nearly every large galaxy has a supermassive black hole at its centre. Supermassive black holes formed in the early universe and have evolved together with galaxies since then. Exactly how such black holes formed is unknown, but their growth must have been rapid, since supermassive black holes have been observed in objects that formed about one billion years after the big bang. Quasars are extremely bright and distant objects in the universe, and their energy source was a mystery. Early theories proposed that supermassive black holes (SMBHs) were the powerhouses behind quasars. These SMBHs can convert a significant portion of the mass they accrete into energy, leading to the immense luminosity observed in quasars.

## 1.2 Active Galactic Nuclei

Active Galactic Nuclei (AGN) are extraordinary astronomical phenomena characterized by exceptionally luminous galactic cores. These intense energy sources are powered by supermassive black holes (SMBHs) at the centers of galaxies, which accrete surrounding matter. As material falls into the SMBH, it releases enormous amounts of energy, making AGN visible across vast cosmic distances. Some AGN produce relativistic jets, streams of particles accelerated to near-light speeds, which can extend far beyond their host galaxies. These jets play a crucial role in shaping galactic environments and intergalactic space. Quasars, a type of AGN, are among the most luminous objects in the universe. Typically found in the early cosmos, quasars are believed to be powered by young, rapidly accreting SMBHs.

Their extreme brightness allows them to outshine their entire host galaxies, making them valuable probes of the distant universe.

### 1.3 Super Massive Black Hole Binaries

A supermassive black hole binary (SMBHB) is a system consisting of two black holes in close orbit around each other, believed to form during galaxy mergers. The life cycle of an SMBHB can be divided into three main stages: inspiral, merger, and ringdown. These stages provide a comprehensive picture of the evolution of supermassive black hole binaries, from their formation to their final stable state, marked by the emission of gravitational waves detectable on Earth.

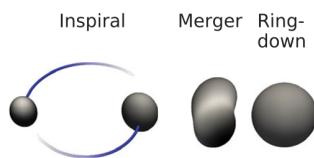


Figure 1.1: A cartoon of the three stages of black-hole coalescence

#### 1.3.1 Inspiral

The first stage, called the inspiral, involves a gradually shrinking orbit. This stage takes a very long time because the gravitational waves emitted are very weak when the black holes are distant from each other. In addition to the orbit shrinking due to the emission of gravitational waves, extra angular momentum may be lost due to interactions with other matter present, such as stars. As the black holes' orbit shrinks, their speed increases, leading to stronger gravitational wave emission. When the black holes are close, the gravitational waves cause the orbit to shrink rapidly.

#### 1.3.2 Merger

The second stage is the merger, characterized by a plunging orbit where the two black holes meet and merge. Gravitational wave emission peaks during this time. The merger is a brief but highly energetic event, producing a burst of gravitational waves detectable by observatories.

#### 1.3.3 Ringdown

The third stage is the ringdown, immediately following the merger. During the ringdown, the newly formed black hole settles into a stable state, emitting gravitational waves as it "rings" like a bell. This ringing is damped over time by the emission of gravitational waves. The ringdown phase starts when the black holes approach each other within the photon sphere. In this region, most of the emitted gravitational waves go towards the event horizon, and the amplitude of those escaping reduces. Remotely detected gravitational waves have an oscillation with a fast-reducing amplitude, as echoes of the merger event result from tighter and tighter spirals around the resulting black hole.

## 1.4 Observational Techniques

Detecting supermassive black hole binaries (SMBHBs) involves various observational techniques across multiple wavelengths and methods. Here are the primary detection methods.

### 1.4.1 Gravitational Wave Observatories

**Space-Based Detectors:** Future missions like LISA (Laser Interferometer Space Antenna) will be specifically designed to detect low-frequency gravitational waves emitted by SMBHBs. LISA will be more sensitive to the inspiral and merger phases of SMBHBs at cosmological distances.

**Pulsar Timing Arrays (PTAs):** PTAs use highly precise measurements of pulsar signals to detect the influence of gravitational waves passing between the pulsars and Earth. These gravitational waves can be generated by SMBHBs. Current PTA projects include Indian Pulsar Timing Array(InPTA) NANOGrav (North American Nanohertz Observatory for Gravitational Waves), the European Pulsar Timing Array (EPTA), Chinese Pulsar Timing Array(CPTA) and Parkes Pulsar Timing Array (PPTA).

### 1.4.2 Electromagnetic Observations

**X-ray Observations:** SMBHBs can emit strong X-rays, particularly if they are accreting gas. The Advanced X-ray Imaging Satellite (AXIS) is a Probe-class concept that will build on the legacy of the Chandra X-ray Observatory by providing low-background, arcsecond-resolution imaging in the 0.3-10 keV band across a 450 arcminute field of view, with an order of magnitude improvement in sensitivity (Reynolds et al. 2023(14)). XMM-Newton can also detect these emissions.

**Radio Observations:** Radio telescopes can identify periodic variations in the emissions from active galactic nuclei (AGN) which might indicate the presence of an SMBHB. Arrays like the Very Large Array (VLA) and the Atacama Large Millimeter-/submillimeter Array (ALMA) are instrumental in such observations.

**Optical and Infrared Observations:** Variability in the light from quasars and active galactic nuclei (AGNs), observed by space-based and ground-based telescopes, can suggest the presence of SMBHBs. Periodic changes in the luminosity or spectral lines can be key indicators. The presence of SMBHBs can be inferred by analyzing Doppler shifts in the spectral lines of AGNs or quasars. Additionally, the periodic variability in the light curves of these objects, caused by the orbital motion of the binary, serves as a telltale sign of SMBHBs.

## 1.5 Objectives

The main objectives of this paper are to find methods to effectively identify SMBHB candidates using time-domain photometric data available now and to simplify and streamline this process for future surveys as well. Trying to also find the best way simulate data that accurately mimics real data we may collect in the future, we attempt to understand the benefits and shortcomings of methods being tested in this project to understand what works best.



## 2. Methods

### 2.1 Simulating Data

In this paper, we have aimed to first simulate light curve data of future and current photometric time-domain surveys to test out the various methods of finding periodicity.

#### 2.1.1 Damped Random Walk

To simulate light curves of quasars, we use a method of modelling the time variability of quasars as a stochastic process between two points in time. This process is a Damped Random Walk (specifically an Ornstein-Uhlenbeck process), (MacLeod et al. 2010(7); Kelly et al. 2009(2); Zu et al. 2013(13); Kozlowski et al. 2010(4)). To simulate light curves using the damped random walk model, we use the PSD (Power Spectral Density) function given by Zu et al. 2013(13);

$$P(f) = \frac{4\sigma^2\tau}{1 + (2\pi\tau f)^2} \quad (2.1)$$

where  $\sigma^2$  is the variance of the light curve data,  $\tau$  is a characteristic DRW time-scale, and  $f$  is the Fourier space frequency. Using the prescription given by Timmer and Koenig (1995)(10) and following the methods of Witt et al. (2022)(11) we take the inverse Fourier transform of the PSD and simulate the data. We have tried to simulate data of current and upcoming surveys like Legacy Survey of Space and Time (LSST) by the upcoming Vera C. Rubin Observatory and the Catalina Real-time Transient Survey (CRTS). Using parameters of different surveys such as the baseline, seasonal gap, cadence, and cadence deviation, we can generate data that mimics the data of each survey. These survey parameters can be added to the program in such a way that in the future, any surveys that might need to be added can be done with ease. Factors such as  $\sigma^2$  and  $\tau$  are also taken randomly

from set ranges to get best possible simulations. For  $\tau$ , we follow Koslowski (2017(5)) and Witt (2022(11)) which takes values in the range  $\tau = [10^{-3}T, 15T]$ , where  $T$  is the observation baseline. As for sigma we take values from log-uniform distribution ranging from [-1.60, -0.25] (Witt et al. 2022(11), MacLeod et al. 2010(7)).

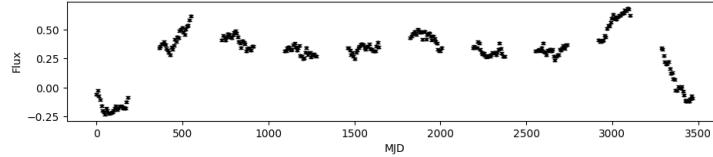


Figure 2.1: Simulated DRW Light Curve using proposed LSST parameters

Other than this method we also used the `astroML` python package which has a function `generate_damped_RW`, which also produced results similar to the inverse Fourier transform method with minor changes such as tweaks in the PSD function.

### 2.1.2 Injecting a sinusoid

To simulate light curves that include SMBHB signals, we inject a sinusoid into the DRW signal (Witt et al. 2022(11)). Adding photometric errors and random gaussian errors to make it as realistic as possible, we get light curves without SMBHB signals (Figure 2.1), as well as with SMBHB signals (Figure 2.2). As for the parameters of the sinusoid, we have the period  $P$ , amplitude  $A$ , and phase  $\phi$ . Given by,

$$x(t) = A \sin \left[ \left( \frac{2\pi}{P}(t - t_0) \right) + \phi \right] \quad (2.2)$$

where  $t_0$  is a reference time. Range of period of the sinusoid is defined by the parameters of the LSST, which is 30 days to 10 years, so that at least one cycle is completed when using LSST parameters (Witt et al. 2022(11)). The phase as well as amplitude are also taken from ranges corresponding to Witt (2022(11)) as  $[0, 2\pi]$  and  $[0.05, 0.5]$  respectively.

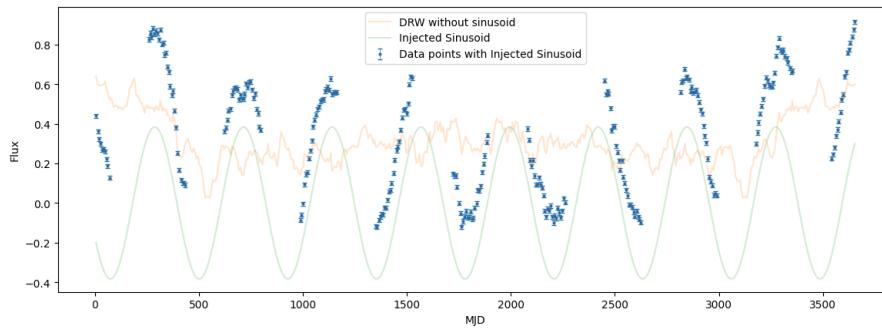


Figure 2.2: Simulated DRW Light Curve with Injected Sinusoid using proposed LSST parameters

## 2.2 Analysis

In this section, we analyze the data we previously simulated. By analyzing it, we identify the periodicity of the quasars. We employ three different methods to determine this periodicity. The three methods we use here are,

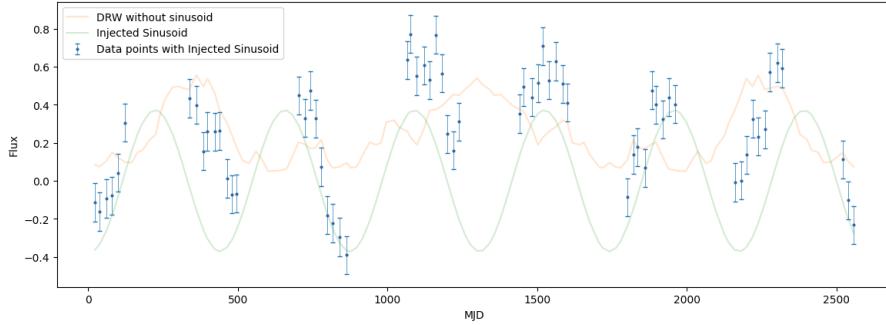


Figure 2.3: Simulated DRW Light Curve with Injected Sinusoid using CRTS parameters

### 2.2.1 Generalized Lomb-Scargle Periodogram

The Lomb-Scargle periodogram (LSP) method iteratively fits an unevenly sampled light curve with sinusoids of various frequencies, generating a periodogram based on the goodness of fit (Lomb 1976(6); Scargle 1982(8)). The Generalized Lomb-Scargle Periodogram (GLSP) is improved from the classical LSP by considering the errors in flux measurements within the light curve. Additionally, the GLSP uses a fitting function that includes both a sinusoid and a constant term (Zechmeister & Kürster 2009(12)). Here we use GLSP code from the PYASTRONOMY python package (Czesla et al. 2019(1)). This code also estimates the significances of the periodogram peaks in terms of the False Alarm Probability (FAP), considering an underlying white-noise periodogram. The GLSP is not effective at detecting transient periodicities in a light curve because the aperiodic components diminish the quality of the sinusoid fit. However, it is highly efficient for identifying persistent periodicities.

### 2.2.2 Weighted wavelet Z-transform

The weighted wavelet Z-transform (WWZ) method is a crucial tool for studying transient periodicities. It can detect the power of any dominant periodic modulation and its corresponding time span in the light curve. This method creates a WWZ map by convolving the light curve with a time- and frequency-dependent kernel, then decomposing the data into time and frequency domains. We convolved the light curve with the Morlet kernel (Grossmann & Morlet 1984) given as

$$f[\omega(t - \tau)] = \exp[i\omega(t - \tau) - c\omega^2(t - \tau)^2] \quad (2.3)$$

The WWZ map is given as

$$W[\omega, \tau; x(t)] = \omega^{1/2} \int x(t) f^*[\omega(t - \tau)] dt \quad (2.4)$$

where  $x(t)$  is the light curve,  $f^*$  is the complex conjugate of the Morlet kernel  $f$ ,  $\omega$  is the frequency, and  $\tau$  is the time-shift parameter. Also WWZ transform is useful when the signal frequency is changing in time.

### 2.2.3 REDFIT

The REDFIT software is a FORTRAN 90 program. This program fits a firstorder autoregressive (AR1) process, to an unevenly spaced time series and has the advantage of avoiding any interpolation in the time domain (Schulz & Mudelsee 2002(9)).

The fitted AR1 model is then transformed from the time domain to the frequency domain to estimate the underlying red-noise spectrum, which characterizes the variability of most blazar emissions. Comparing the spectrum of the actual time series with the estimated red-noise spectrum allows for testing the hypothesis that the origin of the light curve is an AR1 process. By comparing the periodogram of the original time series with those of the synthetic series, the method calculates significance levels for each peak in the periodogram of the original data. Peaks that exceed the significance threshold are considered likely to represent true periodic signals rather than noise. REDFIT is specifically designed to account for red noise, making it more effective than traditional methods (like the Lomb-Scargle periodogram) in distinguishing true periodic signals from noise in datasets where red noise is present.

#### 2.2.4 Bayesian Inference

We use Bayesian model selection to get the AGN light curves. The Bayes' theorem states that

$$P(\theta | \mathbf{d}, H) = \frac{L(d | \theta, H)P(\theta | H)}{Z(\mathbf{d} | H)} \quad (2.5)$$

The posterior probability distribution  $P(\theta | d, H)$  is the probability of the parameters  $\theta$  given the hypothesis  $H$ , before considering the data  $d$ . It reflects our prior belief about the parameters. The likelihood  $L(d | \theta, H)$  is the probability of the data  $d$  given the parameters  $\theta$  and the hypothesis  $H$ . It reflects how likely the observed data is under different parameter values. The likelihood function for the quasar light curves is given by

$$L(d | \theta_n, \theta_s, m) = \frac{1}{\sqrt{(2\pi)^N |C|}} \exp \left[ -\frac{1}{2}(d - m - s)^T C^{-1} (d - m - s) \right] \quad (2.6)$$

The prior Probability  $P(\theta | H)$  is the probability of the parameters  $\theta$  given the hypothesis  $H$ , before considering the data  $d$ . It reflects our prior belief about the parameters. The Bayesian evidence or marginal likelihood for the hypothesis  $H$  is given by

In this work, we utilize measurements of optical magnitudes, represented as the logarithmic light curve. The data  $d$  is modeled as

$$d = n + m + s \quad (2.7)$$

where  $n$  is the noise vector, which encompasses measurement uncertainties and additional intrinsic stochastic variability from the quasar.  $m$  is a constant vector with identical entries, representing the mean magnitude and any constant offset, such as the constant level of contamination from the host galaxy light. The signal vector  $s$  is given by

$$s(t) = A \sin(2\pi f_0 t + \phi) \quad (2.8)$$

where  $A$  is the amplitude,  $f_0$  is the frequency of the sinusoidal signal, and  $\phi$  is the phase (17)

## 3. Results

### 3.1 Observational Data

We analyzed the lightcurve of object PG1302. We used CRTS survey data from Zhu (18). The observations span a period of 2946 days. There are 290 observations during this period. This means average cadence of about 10 days. The magnitude of the object varies between 14.79 and 15.19. A mean photometric error of 0.06 is associated with the observations. The graph of magnitude is plotted against time in 3.1.

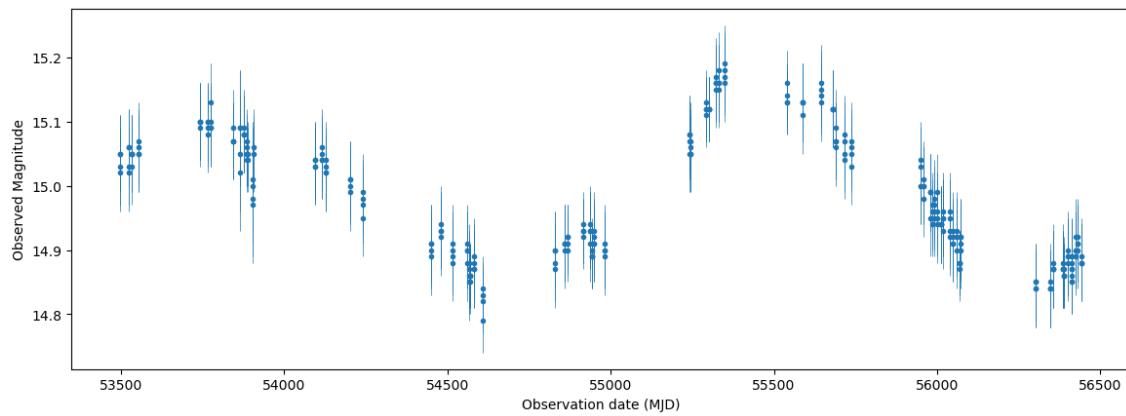


Figure 3.1: PG1302: Observed Lightcurve

### 3.2 Lomb Scargle Periodogram

We created a Generalized Lomb Scargle Periodogram using the Timing module of PyAstronomy ((16)). We used the default options of the module for the analysis. Summary of the analysis is as follows:

Table 3.1: Results of Generalized Lomb Scargle Analysis

Inputs			
Number of input points	290		
Weighted mean of magnitude	14.9804		
Weighted rms of dataset	0.0979		
Time base (days)	2945.97		
Number of frequency points	1449		
Best-Fit Sinusoid			
Maximum power (ZK)	0.8939		
RMS of residuals	0.0319		
Mean weighted internal error	0.0580		
Best sine frequency (cycles/day)	$0.000577 \pm 0.000007$		
Best sine period (days)	$1732.92 \pm 22.00$		
Amplitude	$0.1243 \pm 0.0026$		
Phase (ph) (rad)	$0.0989 \pm 0.0034$		
Phase (T0) (days)	$53324.88 \pm 5.88$		
Offset	$14.9887 \pm 0.0019$		
Top Frequencies			
Rank	Strength	Period (days)	FAP
1	0.8939	1733	$2.3e-138$
2	0.5237	301	$8.5e-45$
3	0.3525	468	$1.2e-25$

In addition to the default options for GLSP, we tried a few variations.

1. Change the ofac (Oversampling factor of frequency grid) to 100 from default value of 100
2. Specify frequency bins to 3000 bins between 1/3000 (low) and 1/30 (high) cycles per day.

The results of the analysis did not change significantly with these variations.

We also investigated how well we could recover the period and amplitude of simulated light curves under different conditions. We focused on the relationship between sigma (the variance term in the PSD function) and the amplitude of the injected sinusoid. Sigma effectively determines how spread out the data points are in the light curve.

We designed a series of simulations where we fixed sigma and varied the amplitude through 5 points in our chosen range. This process was repeated for various sigma values. Our goal was to identify any correlations between these parameters and the accuracy of period and amplitude recovery.

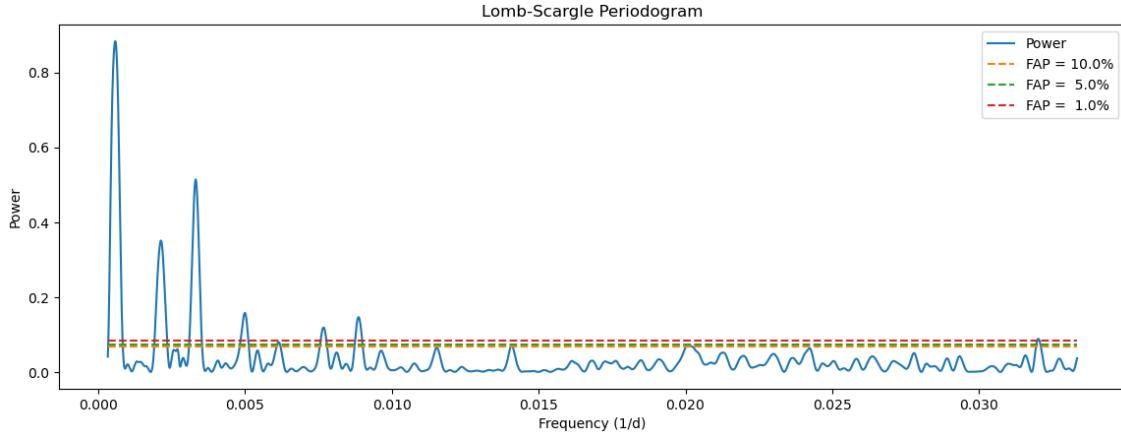


Figure 3.2: PG1302: LS Periodogram (Frequency Domain)

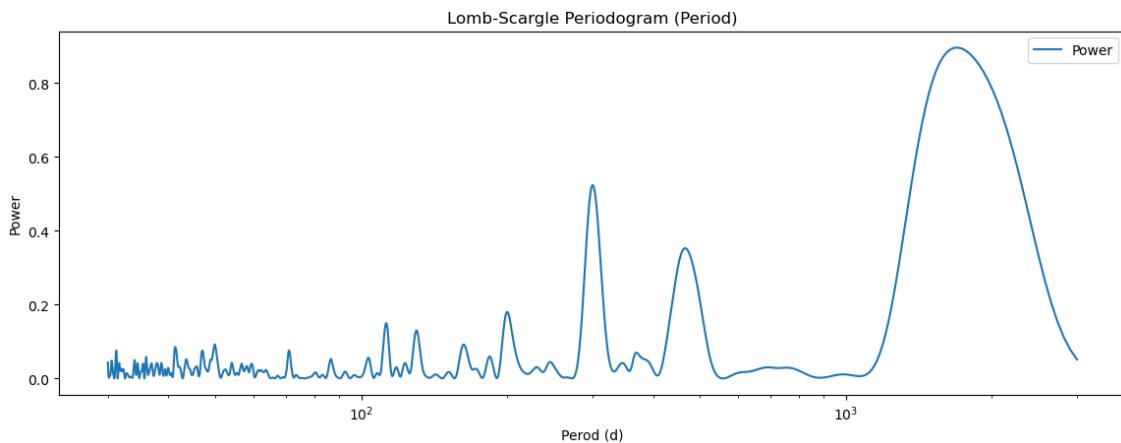


Figure 3.3: PG1302: LS Periodogram (Period Domain)

Upon plotting the results (Figure 3.5), we observed a general trend: period recovery was more accurate for lower sigma values and higher amplitudes. This aligns with expectations, as lower sigma means less spread in the data points, making the periodic signal easier to detect. Similarly, higher amplitudes make the sinusoidal pattern more prominent against any background noise.

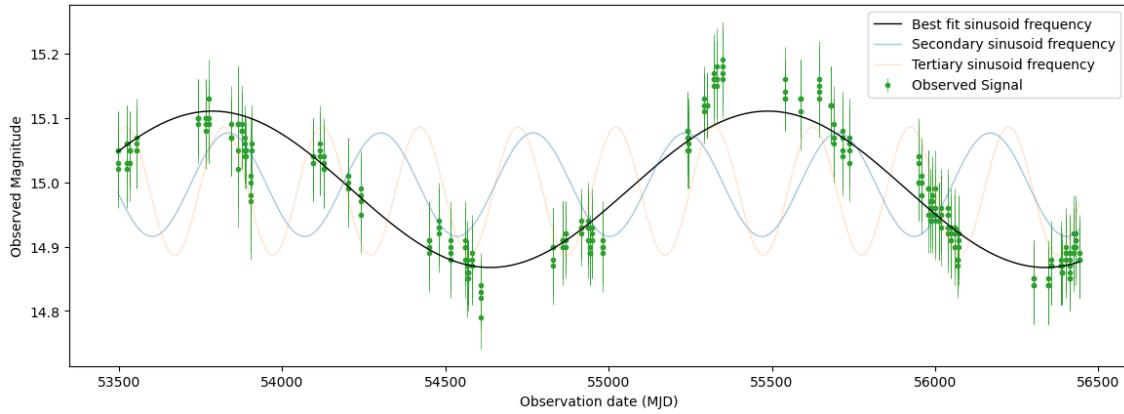


Figure 3.4: PG1302: LSP Fitted Model

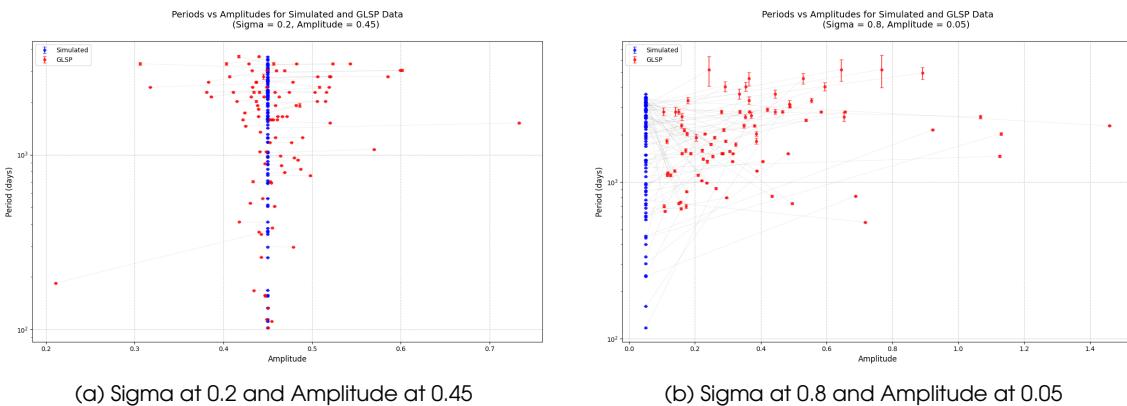


Figure 3.5: Period and Amplitude recovery by GLSP

### 3.3 WWZ Results

We used Weighted Wavelet Z-transform code from Kiehlman et al. (15) We created 1000 frequency bins between 1/3000 (lowest) and 1/30 (highest) cycles-per-day. We used the default bins for Tau (period) parameter. The algorithm created 368 Tau bins based on the data. Result of analysis for the strongest frequency are tabulated below.

Table 3.2: Results of WWZ Analysis

Inputs	
Number of input points	290
Number of frequency bins	1000
Number of Tau bins	368
Results	
Strongest frequency power	0.5824
Strongest frequency(cycle/days)	0.0006
Strongest period(days)	1673
Tau at which strongest frequency was detected	56369.9308

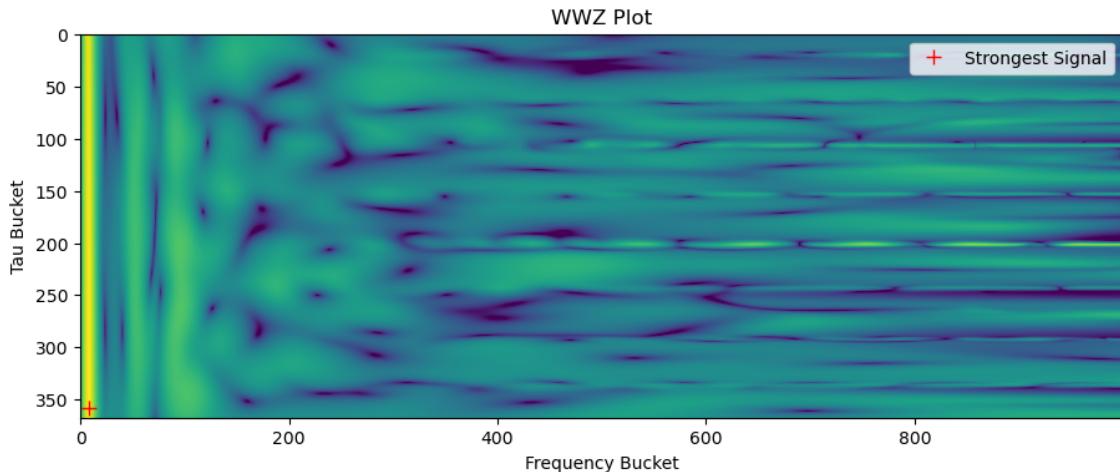


Figure 3.6: PG1302: WWZ - Transform

**Note:** WWZ analysis algorithm produces a warning.

"WARNING: pmax is larger than the maximum time range of the data divided by 5.0. pmax should not be larger than 589.0."

This means, ideally the observation period should be long enough to contain 5 cycles of periodic data.

### 3.4 REDFIT Results

To run the REDFIT program (9), we transformed the CRTS data into a format that REDFIT program expects. The output of REDFIT program, which is written in the form of a ".red" file, is then analyzed and plotted.

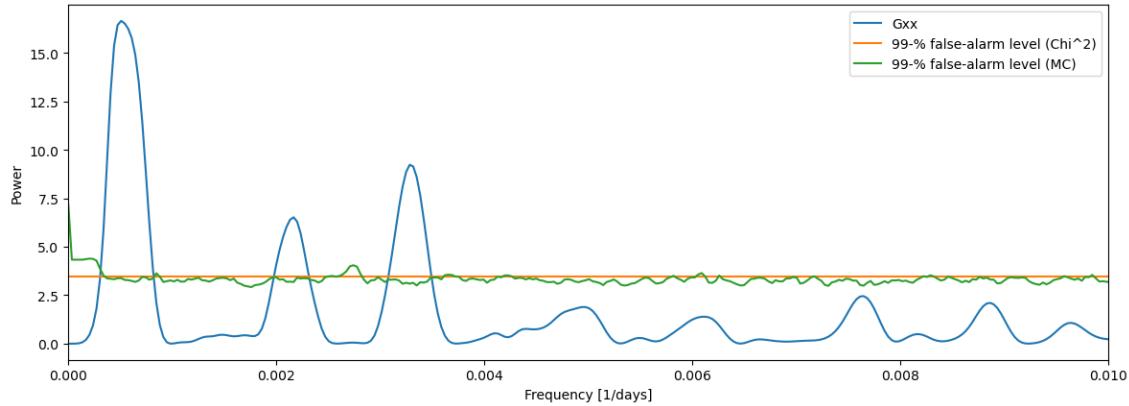


Figure 3.7: PG1302: REDFIT Plot with FAP(False-Alarm Probabilities)

Table 3.3: Results of REDFIT Analysis

Configuration				
Parameter	Description	Default	Configured Value	
nsim	Number of simulations	1000	2000	
mctest	Toggle calculation of False-Alarm Levels based on Monte-Carlo simulation	False	True	
rhopre	Prescribed value for rho; unused if < 0	-99.0	-99.0	
ofac	Oversampling factor for Lomb-Scargle Fourier transform	1.0	10.0	
hifac	Max. frequency to analyze	1.0	1.0	
n50	Number of WOSA segments (with 50 % overlap)	1	1	
iwin	Window-type identifier used to suppress sidelobes in spectral analysis	0: Rectangular	0: Rectangular	
Top Frequencies				
Rank	Period (days)	Power	99% FAP Power level	
1	1971	16.6531	3.4	
2	462	6.5204	3.4	
3	305	9.2295	3.3	

### 3.5 Bayesian Inference Results

We applied Bayesian Inference analysis to two light curves: one from the quasar PG1302 and another from a simulated LSST dataset. This allowed us to compare our method's performance on both real and simulated data. Our analysis focused on recovering several key parameters:

1. Amplitude, phase, and period of the sinusoid
2. Red noise characteristics (variance and damping timescale)
3. Deviation from the standard Damped Random Walk (DRW) model
4. Scale factor for measurement uncertainties
5. Mean magnitude

For each analysis, we generated corner plots to visualize our results. These plots show:

- Posterior distributions for each individual parameter
- 2D distributions illustrating relationships between parameter pairs

The corner plots allow us to assess the quality of parameter recovery, associated uncertainties, and any correlations between parameters. By comparing results from PG1302 and the LSST simulation, we can evaluate our method's performance on real data versus controlled simulations. This Bayesian approach provides a robust framework for characterizing complex light curves, accounting for both periodic signals and various forms of noise typical in astronomical time series data.

#### 3.5.1 Bayes Analysis of PG1302 CRTS Observations

Table 3.4: Bayesian Inference analysis on PG1302 CRTS lightcurve

Parameter	Description	Prior	Inferred
A	Amplitude of Sinusoid	0 to 0.5	0.13
$\phi$	Phase of Sinusoid	0 to $2\pi$	3.33
$T_0$	Period of Sinusoid (years)	0 to 10	4.60
$\log \sigma^2$	Red noise variance on short timescales (DRW)	-6.0 to 0	-3.10
$\log \tau_0$	Red noise damping timescale (DRW)	0.0 to 2.0	0.56
$\gamma$	Deviation of red noise from DRW model	-4.0 to +4.0	0.90
$v$	Scale factor to quantify the over/underestimation of measurement uncertainties	0.1 to 2.0	0.96
m	Mean magnitude plus any constant offset	14.5 to 15.5	14.99

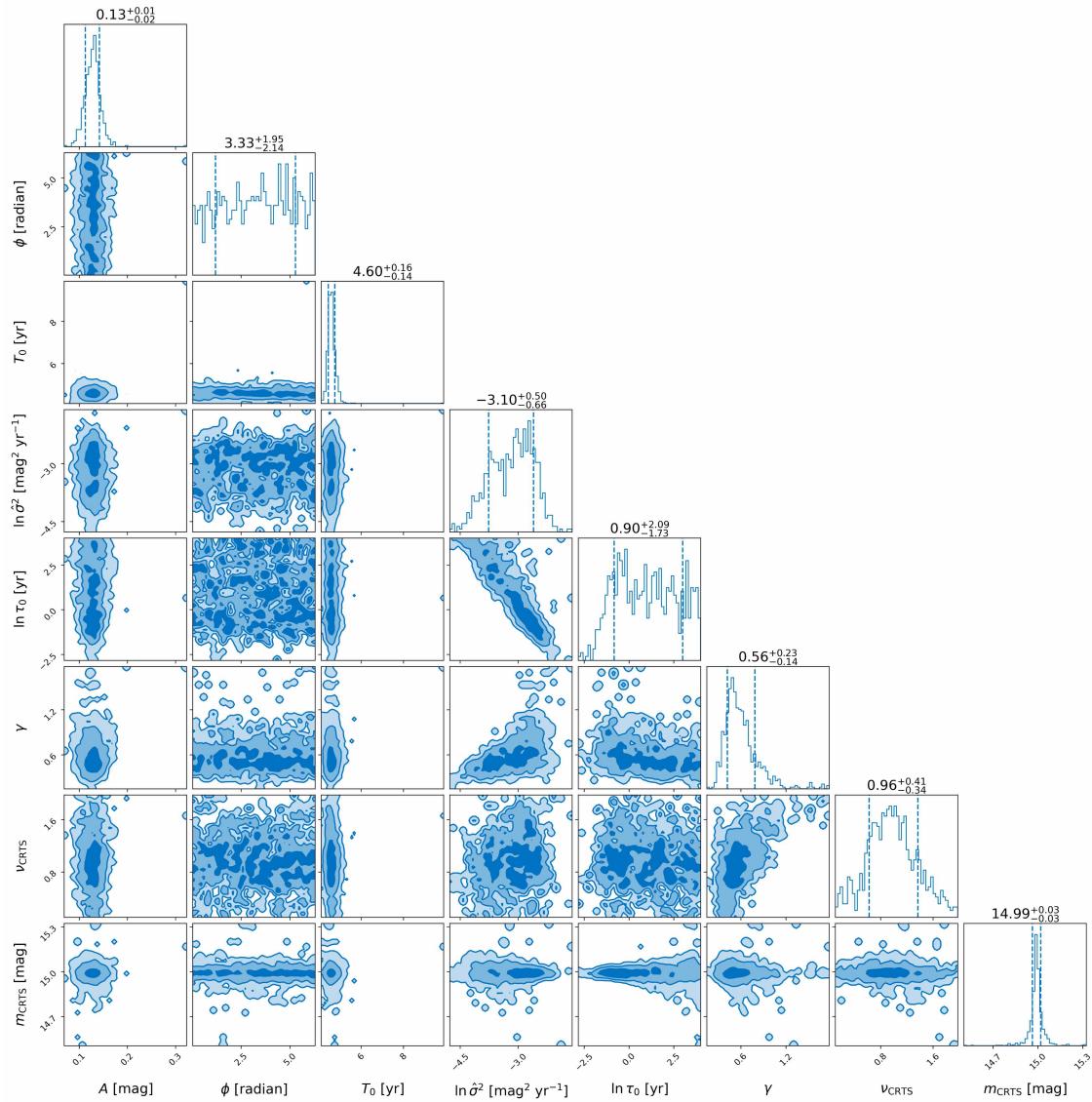


Figure 3.8: Visualizing results of Bayes inference of PG1302 lightcurve

### 3.5.2 Bayes Analysis of Simulated LSST light curve data for a SMBHB

Input light curve was a simulated LSST light curve for a SMBHB. The following graph shows the simulated signal.

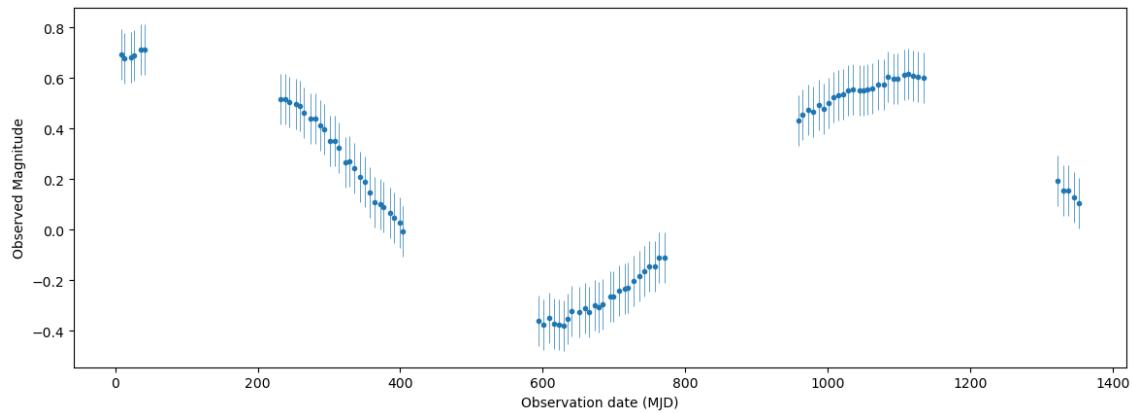


Figure 3.9: Simulated SMBHB Light Curve

Table 3.5: Results of Bayesian Inference on Simulated LSST lightcurve

Parameter	Description	Actual	Prior	Inferred
A	Amplitude of Sinusoid	0.5	0 to 0.6	0.41
$\phi$	Phase of Sinusoid	1.0	0 to $2\pi$	0.98
$T_0$	Period of Sinusoid (years)	2.74	0 to 10	2.67
$\log \sigma^2$	Red noise variance on short timescales (DRW)	-2.0	-6.0 to 0	-1.09
$\log \tau_0$	Red noise damping timescale (DRW)	4.40	-4.0 to +4.0	3.98

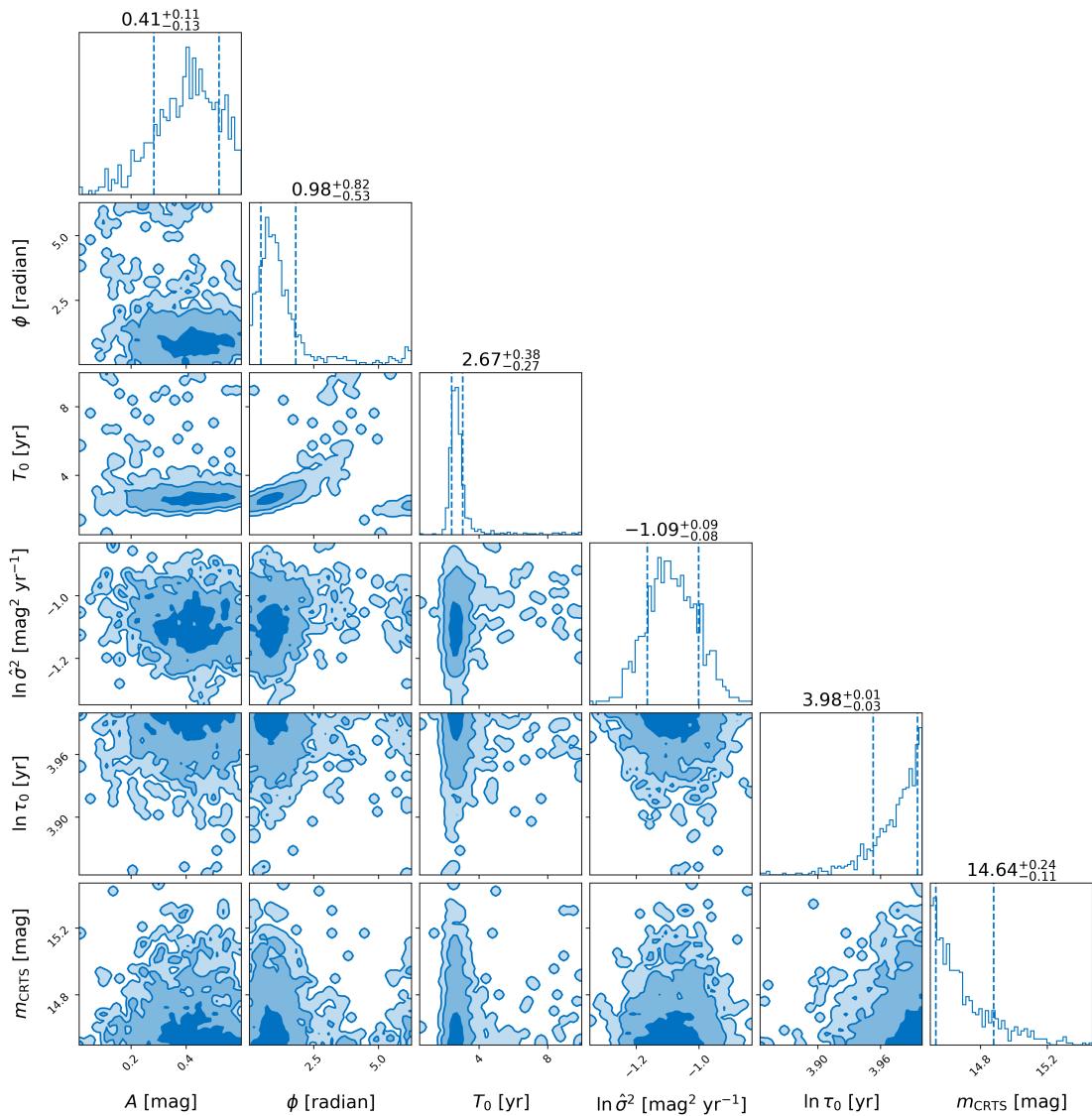


Figure 3.10: Bayes Analysis: Corner plot - all parameters

## 4. Conclusion

### 4.1 Comparison of Results

#### 4.1.1 Accuracy

REDFIT program (9) gave an estimate of periodicity for PG1302 as 1971 days. This result matches very well the result ( $1972 \pm 254$  days) from Kovačević et al (3) GLSP analysis estimated the period as 1733 days. WWZ analysis estimated the period as 1673 days.

#### 4.1.2 Time taken for computation

Below time observations are just for a comparative study of the different methods of analysis. These are taken on a development Windows laptop with 2.2GHz intel i7 processor with 16GB RAM. These are not standardized performance tests.

1. GLSP: 60 ms (Gls function execution time within Python script)
2. REDFIT: 6 sec (Commandline execution time, includes time to read and write files)
3. WWZ: 55.5 sec (WWZ.transform function execution time within Python script, for 1000 frequency X 368 tau bins. The execution time increases with more bins)
4. Bayesian Inference: 1 hour to run the analysis with 72 observation data points and 1000 live points, The performance analysis shows that Bayesian Inference is computationally very expensive. It is very thorough as it provides inference for all the parameters of the light curve. Other methods provide estimation for the periodicity in the light curve. Among these methods, WWZ transform takes more computation time due to two-dimensional nature of analysis.

#### 4.1.3 Applicability

WWZ analysis is relevant for signals in which the periodicity changes with time. It expects that ideally the observation should be long enough to include more than 5 cycles.

Bayesian analysis allows us to create a complex model with multiple parameters and infer those. Other techniques we explored allowed us to estimate the periodicity and amplitude.

## **4.2 Summary**

The primary aim of this research was to determine the periodicity of Active Galactic Nuclei (AGN) using various analytical techniques. By examining Supermassive Black Hole Binaries (SMBHB), AGN, gravitational waves, and Pulsar Timing Arrays (PTA), we provided a comprehensive introduction to the fundamental concepts of our study.

To simulate the light curves, we employed the Damped Random Walk (DRW) process and introduced sinusoidal signals to the data, creating a robust foundation for our subsequent analysis. Our investigation encompassed several methods for identifying periodicity, including the Lomb-Scargle periodogram, Wavelet Weighted Z-transform (WWZ), REDFIT, and Bayesian inference.

Our results demonstrated that among these techniques, Bayesian inference, while valuable, proved to be computationally intensive, suggesting a trade-off between precision and resource expenditure.

This study also underscores the necessity for continued exploration of computationally efficient methods. Future research could further optimize these techniques or develop novel approaches, ultimately enhancing our understanding of SMBHB and their associated phenomena.



## 5. Bibliography

### Books

### Articles

- (Cze+19a) S Czesla et al. "Astrophysics Source Code Library, record ascl: 1906.010". In: (2019) (cited on page 11).
- (KBS09) Brandon C. Kelly, Jill Bechtold, and Aneta Siemiginowska. "ARE THE VARIATIONS IN QUASAR OPTICAL FLUX DRIVEN BY THERMAL FLUCTUATIONS?" In: *The Astrophysical Journal* 698.1 (May 2009), page 895. DOI: 10.1088/0004-637X/698/1/895. URL: <https://dx.doi.org/10.1088/0004-637X/698/1/895> (cited on page 9).
- (Kov+19) Andjelka B. Kovačević et al. "The Optical Variability of Supermassive Black Hole Binary Candidate PG 1302–102: Periodicity and Perturbation in the Light Curve". In: *The Astrophysical Journal* 871.1 (Jan. 2019), page 32. DOI: 10.3847/1538-4357/aaf731. URL: <https://dx.doi.org/10.3847/1538-4357/aaf731> (cited on page 23).
- (Koz+10) Szymon Kozłowski et al. "Quantifying Quasar Variability as Part of a General Approach to Classifying Continuously Varying Sources". In: *ApJ* 708.2 (Jan. 2010), pages 927–945. DOI: 10.1088/0004-637X/708/2/927. arXiv: 0909.1326 [astro-ph.CO] (cited on page 9).
- (Koz17) Kozłowski, Szymon. "Limitations on the recovery of the true AGN variability parameters using damped random walk modeling". In: *AA* 597 (2017), A128. DOI: 10.1051/0004-6361/201629890. URL: <https://doi.org/10.1051/0004-6361/201629890> (cited on page 10).
- (Lom76) Nicholas R. Lomb. "Least-squares frequency analysis of unequally spaced data". In: *Astrophysics and Space Science* 39 (1976), pages 447–462. URL: <https://api.semanticscholar.org/CorpusID:2671466> (cited on page 11).

- (Mac+10) C. L. MacLeod et al. “MODELING THE TIME VARIABILITY OF SDSS STRIPE 82 QUASARS AS A DAMPED RANDOM WALK”. In: *The Astrophysical Journal* 721.2 (Sept. 2010), page 1014. DOI: 10.1088/0004-637X/721/2/1014. URL: <https://dx.doi.org/10.1088/0004-637X/721/2/1014> (cited on pages 9, 10).
- (Sca82) J. D. Scargle. “Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data.” In: *ApJ* 263 (Dec. 1982), pages 835–853. DOI: 10.1086/160554 (cited on page 11).
- (SM02) Michael Schulz and Manfred Mudelsee. “REFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series”. In: *Computers & Geosciences* 28.3 (2002), pages 421–426 (cited on pages 11, 17, 23).
- (TK95) J. Timmer and M. König. “On generating power law noise.” In: *A&A* 300 (Aug. 1995), page 707 (cited on page 9).
- (Wit+22) Caitlin A. Witt et al. “Quasars with Periodic Variability: Capabilities and Limitations of Bayesian Searches for Supermassive Black Hole Binaries in Time-domain Surveys”. In: *The Astrophysical Journal* 936.1 (Sept. 2022), page 89. ISSN: 1538-4357. DOI: 10.3847/1538-4357/ac8356. URL: <http://dx.doi.org/10.3847/1538-4357/ac8356> (cited on pages 9, 10).
- (ZK09) Mathias Zechmeister and Martin Kürster. “The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms”. In: *Astronomy and Astrophysics* 496 (2009), pages 577–584. URL: <https://api.semanticscholar.org/CorpusID:10408194> (cited on page 11).
- (Zu+13) Ying Zu et al. “IS QUASAR OPTICAL VARIABILITY A DAMPED RANDOM WALK?” In: *The Astrophysical Journal* 765.2 (Feb. 2013), page 106. DOI: 10.1088/0004-637X/765/2/106. URL: <https://dx.doi.org/10.1088/0004-637X/765/2/106> (cited on page 9).

## Conference Proceedings

- (Rey+23) Christopher Reynolds et al. “Overview of the advanced x-ray imaging satellite (AXIS)”. In: *UV, X-Ray, and Gamma-Ray Space Instrumentation for Astronomy XXIII*. Edited by Oswald H. Siegmund and Keri Hoadley. SPIE, Oct. 2023. DOI: 10.1117/12.2677468. URL: <http://dx.doi.org/10.1117/12.2677468> (cited on page 7).

## Software

- (KMK21) Sebastian Kiehlmann, Walter Max-Moerbeck, and Oliver King. *wwz: Weighted wavelet z-transform code*. Version 0. Sept. 2021 (cited on page 17).

## Miscellaneous

- (Cze+19b) Stefan Czesla et al. *PyA: Python astronomy-related packages*. June 2019. ascl: 1906.010 (cited on page 14).

- (Thr20) Xingjiang Zhu Eric Thrane. *Toward the unambiguous identification of supermassive binary black holes through Bayesian inference*. arXiv preprint arXiv:2004.10944v3 (astro-ph.HE). Aug. 2020 (cited on page 12).
- (Zhu20) Xingjiang Zhu. *SuperBayes*. <https://github.com/ZhuXJ1/SuperBayes/blob/master/data>. 2020 (cited on page 13).

