University of Southampton

Faculty of Faculty of Engineering and Physical Sciences
School of Physics and Astronomy

A data-driven study of variability in X-ray binaries using machine learning

by

Jakub Kacper Orwat-Kapola
ORCiD: 0000-0002-3984-1881

A thesis for the degree of
Doctor of Philosophy

February 2024

University of Southampton

Abstract

Faculty of Faculty of Engineering and Physical Sciences
School of Physics and Astronomy

Doctor of Philosophy

A data-driven study of variability in X-ray binaries using machine learning

by Jakub Kacper Orwat-Kapola

The field of time series data mining has become increasingly important in light of the impending era of "Big Data". With new astronomical sky surveys set to generate data at an unprecedented rate, the need for automated methods of data analysis is becoming increasingly urgent.

This thesis makes a contribution to the toolbox of automated characterisation methods for astronomical time series, specifically light curves. Two end-to-end pipelines drawing on concepts from the fields of signal processing, data science, machine learning, and deep learning are presented, with examples showing how they can streamline the process of light curve analysis.

The first pipeline is designed to extract and aggregate features from light curve segments, using a neural network and clustering algorithms to combine them into feature vectors of fixed length. These "light curve fingerprints" can be used as input for downstream machine learning algorithms, providing a way to tackle the problem of constructing a standard representation for light curves of variable length. I demonstrate the proposed method on the X-ray data of the Galactic low-mass black hole X-ray binary system GRS 1915+105, showcasing how it can be used to study the variability of the source on time scales ranging from seconds to hours, and to quantify the similarity of different light curves. The outcomes highlight the problem of the popular classification system of GRS 1915+105 observations, which does not account for intermediate class behaviour. I also explore the extension to another source, and outline the direction of further development.

The fourth chapter of the thesis presents a periodicity detection and characterisation pipeline, which utilises the wavelet transform for timing-frequency analysis. The pipeline is applied to a sample of photometric light curves of Small Magellanic Cloud Be X-ray binaries, demonstrating its ability to facilitate analysis of the sources' behaviour over time scales ranging from days to decades. Using this methodology, I detect several periodic signals for the first time, and provide the interpretation of their physical origin. The pipeline is a useful tool in the analysis of the binary orbital behaviour of the sources, the evolution of Be star non-radial pulsations, and their possible link with the formation mechanism of the circumstellar decretion discs.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
   Jakub K. Orwat-Kapola, Antony J. Bird, Adam B. Hill, Diego Altamirano, and Daniela Huppenkothen. Light-curve fingerprints: An automated approach to the extraction of X-ray variability patterns with feature aggregation - An example application to GRS 1915+105. *Monthly Notices of the Royal Astronomical Society*, 509(1):1269–1290, 1 2022. ISSN 13652966.

Signed:..... Date:.....

Acknowledgements

Here, I would like to thank everyone who has contributed to this thesis with their feedback, including Daniela Huppenkothen, who provided insightful comments and had a great impact on parts of this work, Sebastian Drave and Poshak Gandhi who examined it, as well as the data scientists of HAL24K Labs, who created a great working environment during my time in Amsterdam.
Many thanks to my supervisors; Diego Altamirano, Tony Bird and Adam Hill, for giving me the opportunity to do this work, and for their mentorship and encouragement along the way.
A massive thank you to Alessandra, Andrés, Arianna, Cory, Hao, Kevin, Maria, Matt, Mayukh, Mehtap, Miika, Noel, Norman, Yanan, and all the friends and colleagues from the Southampton Astronomy Group. These last few years would not be the same without those great people.
Dziękuję Mamo i Tato, za wasze wsparcie.

Definitions and Abbreviations

ANN	Artificial neural network
BeXRB	Be X-ray binary
BH	Black hole
BIC	Bayesian information criterion
CNN	Convolutional neural network
FAP	False alarm probability
FATS	Feature analysis for time series
FRED	Fast rise exponential decay
FWHM	Full width at half-maximum
GMM	Gaussian mixture model
HJD	Heliocentric Julian Date
HMXB	High-mass X-ray binary
IFoS	Intensity features of light curve segments
LMXB	Low-mass X-ray binary
LSST	Large Synoptic Survey Telescope
LSTM	Long-short term memory
MJD	Modified Julian Date
ML	Machine learning
NRP	Non-radial pulsation
NS	Neutron star
OGLE	Optical Gravitational Lensing Experiment
PA	Phase asymmetry
PCU	Proportional Counting Unit of the RXTE/PCA
PS	Phase span
RMS	Root-mean-square
RNN	Recurrent neural network
RXTE	Rossi X-ray Timing Explorer
RXTE/PCA	Proportional Counter Array on-board of the Rossi X-ray Timing Explorer
SFoS	Shape features of light curve segments
SGD	Stochastic gradient descent
SIFoS	Shape and intensity features of light curve segments
SMC	Small Magellanic Cloud

STFT	Short-time Fourier transform
SXP	Small Magellanic Cloud X-ray Pulsar
t-SNE	t-distributed stochastic neighbour embedding
TJD	Truncated Julian Date (Heliocentric Julian Date minus 2.45×10^6 days)
UMAP	Uniform Manifold Approximation and Projection
VAE	Variational autoencoder
WWZ	Weighted wavelet Z-transform
XRB	X-ray binary
XROM	X-ray Variables OGLE Monitoring System
ZTF	Zwicky Transient Facility

Chapter 1

Introduction

This work focuses on the exploration of data-driven analysis methods of variability in X-ray binary systems. Specifically, the application of machine learning (ML) methods to the characterisation of X-ray light curves. The application of non-ML based, signal processing methods to the analysis of X-ray binary photometry is also discussed. This chapter explains why the development of data-driven methodologies is particularly important now, it introduces the physics and variability of X-ray binaries, as well as the methodologies used in later chapters.

Automated approaches to data analysis are becoming increasingly relevant as we are entering the era of big data in and outside of astronomy. For example, the healthcare services generate large amounts of data in the form of electronic medical records, clinical trials, lab results, genomic data etc. (Huser and Cimino, 2016), providing healthcare professionals with an opportunity to gain insights into patient outcomes, treatment effectiveness, and disease patterns. New capabilities for data acquisition and processing with ML are leading to initiatives which aim to develop personalised medicine that mitigates drug resistance issues (Nussinov et al., 2021). Time series data analysis and modelling methods have a wide array of potential uses, including the prediction of cardiac arrest risk (Kennedy and Turley, 2011), intensive care unit readmission risk (Lin et al., 2019), the prediction of patient medication expenditures (Kaushik et al., 2020) or the prediction of hospital-onset COVID-19 infections (Myall et al., 2022). Other potential applications of ML include predictive policing (Shah et al., 2021) and police response analysis (Rotaru et al., 2022), fraud detection (Varmedja et al., 2019) etc. Some specific industrial applications include, for example, analysis of smart utility meter data and city traffic data. The ability to identify appliance energy usage patterns can provide actionable insights to utility providers (Singh and Yassine, 2018), as smart meters are being installed in millions of houses. Prediction of the rate of traffic flow and smart route planning can aid in the management of congestion in intelligent transportation systems (Zhu et al., 2019).

The need for study of large data sets across multiple scientific disciplines propelled development in the areas of bioinformatics (with example applications in genomics, Bayat, 2002), geoinformatics (e.g. navigation, land surveying, Praveen et al., 2016), cheminformatics (e.g.

pharmacology Xia, 2017) and astroinformatics. All of these disciplines combine information technology, data science and domain-specific knowledge to extract useful information from the data. This often involves the application of statistical and ML methods to identify patterns in the data. Various methods of modelling of time series data have been employed in numerous fields of research (e.g. Längkvist et al., 2014; Hyndman et al., 2015; Benkabou et al., 2018; Ismail Fawaz et al., 2019).

Within the field of astronomy, future and ongoing surveys, like those conducted by Vera C. Rubin Observatory (also known as the Large Synoptic Survey Telescope, LSST, Ivezic et al., 2019) and Zwicky Transient Facility (ZTF, Bellm, 2014), will produce petabytes of data at unprecedented rates. LSST is a new synoptic survey currently under construction, with the completion expected in 2024. It will operate for 10 years, imaging the entire visible sky every 3–4 nights, using six optical filters ($ugrizy$). Its highly sensitive camera is expected to discover 20 billion galaxies and a similar number of stars. The ZTF survey covers the observable northern sky every two nights, using g and r filters. The most recent, sixteenth data release contained 4.64 billion light curves. Given that manual analysis of this volume of data by human experts is impossible, we need to develop systems capable of interpreting the data for us.

1.1 Data science for astronomical time series

Light curves are a type of time series data describing the intensity of electromagnetic emission from an astronomical source. They can show regular, semiregular and irregular behaviour, depending on the periodicity of the signal, or lack thereof, which is dictated by the nature and state of the observed object. Figure 1.1 shows the vast range of possible object classifications, arranged according to the causes of their variability (see Eyer et al., 2019, for an overview of variability in the second data release of the Gaia survey). The two main branches of this diagram differentiate objects depending on whether the variability is intrinsic or extrinsic to the object itself. The branches are further divided according to the stellar or non-stellar nature of the source, and then further, according to the specific variability class. Sources showing intrinsic causes of variability include active galactic nuclei, eruptive variables, cataclysmic variables and pulsating variables. Examples of extrinsic causes include rotation, eclipsing and microlensing.

Variability of the sources gives us a lot of information about the physical processes taking place within the source. Solar flares are a type of eruptive phenomena associated with a sudden energy release caused by magnetic reconnection, which for example is observed in the Sun, as well as others stars and young stellar objects (see Benz and Güdel, 2010, for a review). Cataclysmic variables are binary systems consisting of a white dwarf and a low mass, main sequence star, where the white dwarf accretes through Roche lobe overflow. Variability of these sources can be affected by the strength of the white dwarf's magnetic field, and thanks to their relatively short evolution timescales they can serve as laboratories for the study of the accretion physics (Scaringi et al., 2022). Radial and non-radial pulsations are observed in a wide variety of stars

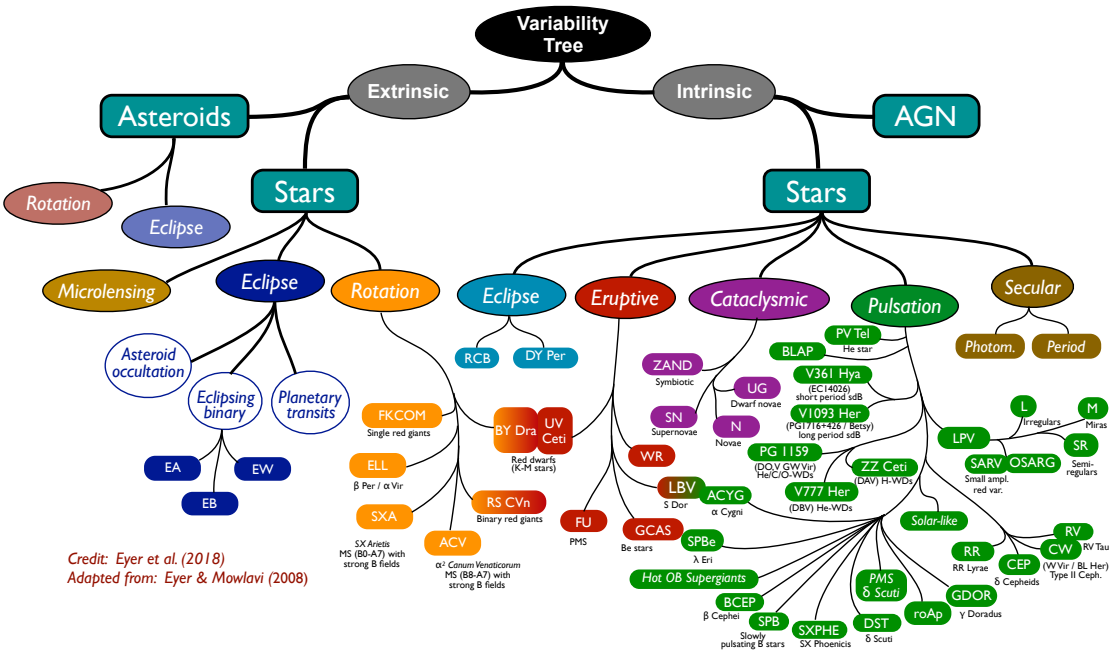


FIGURE 1.1: Variability tree diagram showing the multiplicity of different classes of variable objects observed in astronomy. Credit: Eyer et al. (2019)

with different temperatures and luminosities, scattered across the Hertzsprung–Russell diagram, and are triggered by the partial ionisation of hydrogen and helium. For a more detailed review of different classes of stellar variability, see Catelan and Smith (2015).

In order to study these physical processes, we need to observe the relevant sources over timescales which are adequate for their particular types of variability. The purpose of large astronomical surveys is to provide such observational data, however in order to make sense of this large volume of data and identify relevant observations, analysts need methods which can extract descriptive variables from individual data observations (i.e. perform feature engineering) and use them in ML pipelines to compare observations and perform tasks like classification, outlier detection or clustering.

One example of a project that aims to identify observations of interest using ML is the Automatic Learning for the Rapid Classification of Events (ALERCE). ALERCE is a data broker which is currently processing the ZTF data, and is preparing to become a broker for LSST. ALERCE (Sánchez-Sáez et al., 2021) have developed a light curve classification pipeline, which generates variability features from ZTF light curves, and colours from AllWISE and ZTF photometry (containing information about the relative intensity of different photometric bands), to perform classification using a type of random forest algorithm. Their classifier uses 152 sophisticated light curve features, which for example include values describing whether the periodogram of

the light curve is dominated by one particular frequency, the intensity of peaks potentially caused by aliasing, parameters of an irregularly sampled variant of an autoregressive model, light curve variance estimates for short and long time scales, history of non-detections of the target, as well as a set of other statistical features, generated by a library for time series feature extraction called "feature analysis for time series" (FATS, Nun et al., 2015). This is an example of a classification pipeline that uses a curated set of light curve descriptors developed by domain experts.

ML-based inference is most often preceded by this labour-intensive task of feature engineering, which requires domain-specific knowledge. This approach also starts from the assumption that we know what we are looking for in the data set, and such specific features might fail to capture unexpected, anomalous behaviour. Automated feature extraction methods aim to remove this requirement of interpretation of data using expert knowledge, and to reduce the biases associated with it. Several methods of automated feature extraction for light curve data have been studied in the past, for example Armstrong et al. (2015) used Kohonen self-organising maps to perform the classification of variable star light curves, while Pieringer et al. (2019) used pattern dictionary learning to encode information contained in photometric light curves, and highlight the most salient light curve sections, like peaks and drops in magnitude. Mackenzie et al. (2016) extracted light curve segments and clustered them to find common patterns of variability in variable star candidates, creating a representation compatible with ML classifiers. Similarly, Valenzuela and Pichara (2018) used a sliding window method to extract light curve segments and classified the light curves based on the presence of characteristic patterns. Naul et al. (2018) extracted features of phase-folded light curves using an autoencoding recurrent neural network and used a random forest classifier on observations of variable sources from All Sky Automated Survey Catalog, Lincoln Near-Earth Asteroid Research survey, and Massive Compact Halo Object Project, achieving classification accuracy of well over 90%.

End-to-end architectures have also been used for a variety of tasks. Those approaches use a single model which implicitly learns a low-dimensional representation of data to perform classification, inference etc. For example, Charnock and Moss (2017) used a bidirectional recurrent neural network (RNN) to process and classify light curves of the Supernovae Photometric Classification Challenge data set and achieved impressive results. Mahabal et al. (2017) transformed the light curves of Catalina Real-Time Transient Survey to two-dimensional images and used a convolutional neural network (CNN) to classify them. Shallue and Vanderburg (2018) trained a deep CNN to detect exoplanet transits in folded light curves of the Kepler mission, and were able to discern plausible planet signals from false-positives 98.8% of the time. Becker et al. (2020) trained an RNN to classify variable star light curves, achieving results which were competitive with results of a random forest classifier trained on features generated by FATS. They grouped observations of each light curve with a sliding window, reducing the sequence length, which allowed the RNN to process long sequences. More recently, Zhang and Bloom (2021) demonstrated state-of-the-art classification performance using novel Cyclic-Permutation Invariant Neural Networks which are invariant to phase shifts of phase-folded light curves.

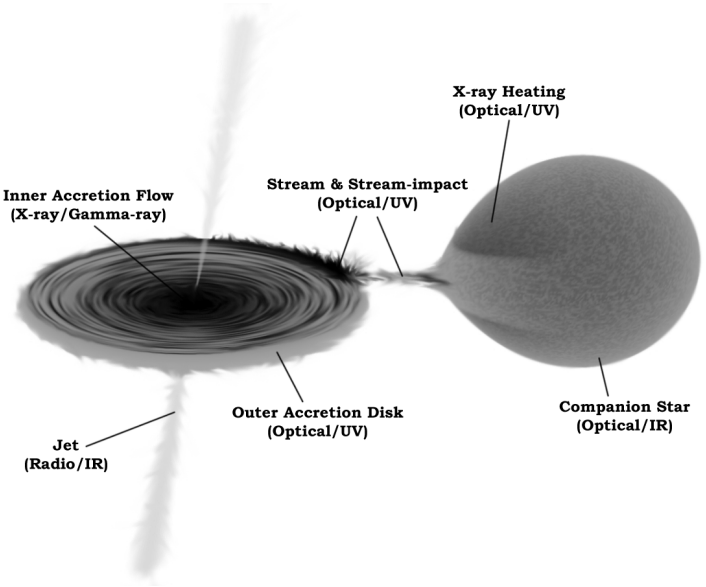


FIGURE 1.2: An illustration of a low mass X-ray binary system. The colour scale is inverted, with the darkest colours representing the brightest regions. Credit: Hynes (2012)

In this thesis, I focus on the application of ML-based methods and signal processing methods to the analysis of variability in light curves of one subset of variable sources; X-ray binary systems. Below, I introduce some fundamental information about these systems, about the physics causing the variability we observe, and discuss the methods of variability analysis and ML in more detail.

1.2 X-ray binaries: causes of variability

X-ray binary systems (XRBs) are comprised of a compact object in the form of a black hole (BH) or a neutron star (NS), and an optical companion star. XRBs are classified as either low-mass (LMXB) or high-mass (HMXB), based on the spectral type of the companion star and the binary mass function. In LMXBs the companions are optical counterparts with the spectral type of A or later, and the mass of $\leq 1 M_{\odot}$, whilst in HMXBs the companion is an early-type, O or B star with the mass of $>10 M_{\odot}$. Systems with companion masses in the range of $1 M_{\odot} \lesssim M \lesssim 10 M_{\odot}$ are referred to as intermediate-mass X-ray binaries.

1.2.1 Low-mass X-ray binaries

Figure 1.2 illustrates an example LMXB, and shows the main types of electromagnetic radiation emitted by the different components of the binary system. These systems have orbital periods ranging from sub-hour to a few weeks. As of 2022, ~200 confirmed and candidate Galactic LMXBs have been identified, and among those, NS systems are twice as abundant as those containing a BH (Bahramian and Degenaar, 2023). The phenomenology of NS LMXBs can differ from that of a BH ones, due to the presence of a solid surface of the NS, and this can

be the basis for identification of the nature of the compact object. These characteristic events include Type I X-ray bursts and pulsations. Type I bursts are thermonuclear flashes caused by rapid fusion of accreted material, which burns accumulated hydrogen into helium and then into carbon, resulting in a flash that reaches maximal intensity in a duration ranging from a fraction of a second to ∼10 s, and then decays over the time span of seconds to minutes (see Lewin et al., 1993, for a review of X-ray bursts). Coherent pulsations are caused by a hot spot on the surface of the spinning NS. Pulsations and Type I X-ray bursts are observed in both LMXBs and HMXBs.

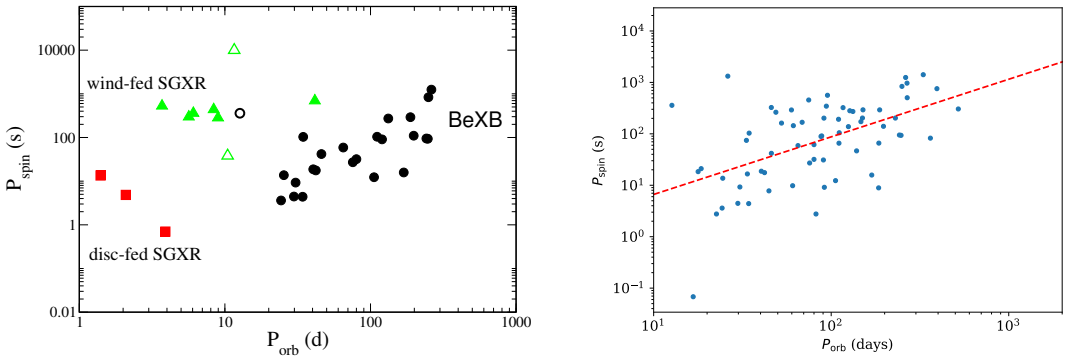
LMXBs can be classified according to their long term X-ray variability, with persistent sources showing stable luminosity levels due to consistent accretion, whilst transient sources can show X-ray luminosity changes by a factor of $\gtrsim$1000, switching between periods of active accretion and quiescence. The majority of LMXBs are X-ray novae, which show significant changes in X-ray luminosity. While transient outbursts typically last for a few weeks up to a ∼year, with recurrence of 1–60 years, a few quasi-persistent sources showing extended outbursts have been observed as well. The active, outburst state in those sources can last for multiple years. Examples of quasi-persistent sources include the NS LMXB EXO 0748-676, which showed an outburst lasting 24 years (Degenaar et al., 2011), the BH LMXB GRS 1915+105, which has been in outburst since 1992 and is still active in X-rays as of February 2023 (Miller and Homan, 2023), the NS system IGR J17062-6143 which has been in the active state since 2006 (Hernández Santisteban et al., 2019) etc. The cause of the outbursts has been modelled as an instability in the accretion disc, caused by changing accretion rate (see Remillard and McClintock, 2006; Bahramian and Degenaar, 2023, for reviews of the properties of LMXBs).

On the high-frequency end of the variability spectrum, many LMXBs exhibit quasi-periodic oscillations (QPO), which have frequencies in the range of 0.01–450 Hz for BH systems, and in NS systems reaching frequencies of up to ∼1200 Hz. These high frequency oscillations are thought to originate from regions very close to the compact object, and are often associated with the inner disc radius. Relativistic precession of the inner hot accretion flow has been proposed as a possible explanation of QPOs in black hole XRBs (see Ingram and Motta, 2019, for a review of QPOs).

In this thesis, I focus on the X-ray variability of two BH LMXBs in the time scales of seconds to hours. Those sources are GRS 1915+105 and IGR J17091−3624, discussed in Chapters 2 and 3 respectively.

1.2.2 High-mass X-ray binaries

HMXBs can be subdivided into two groups; systems containing B-type stars that show optical emission lines (hence often referred to as a Be star), with a spectral type of III-V, and those containing an OB supergiant. Another method of source classification for pulsating HMXB is based on the NS spin and orbital periods, which define a given source's position on the Corbet diagram (Corbet, 1986), as shown in Figure 1.3. Three populations of sources can be



(a) (b)

FIGURE 1.3: (a) Orbital-spin period diagram (Corbet diagram) showing the population of HMXBs. Peculiar, outlier systems are indicated by open markers; the open triangles correspond to 2S0114+65 and OAO 1657−41, while the open circle to SAX J2103.5+4545. (b) Linear fit to the BeXRB sources, given by Equation (1.1). Credit: Reig (2011); Igoshev et al. (2021)

distinguished in the Corbet diagram; Be X-ray Binaries (BeXRB) have relatively long orbital periods of $\gtrsim$15 days that show a correlation with the spin periods, whilst disc-fed supergiant XRBs with a short spin period of $\sim 10^0 - 10^1$ s show an anti-correlation, and wind-fed supergiant XRBs with longer spin periods $\sim 10^2 - 10^3$ s show no correlation.

In this work I focus on BeXRBs, HMXBs containing a Be star and a NS (see Reig, 2011, for a review of BeXRBs). Figure 1.3b shows the Corbet diagram of the BeXRB sources found in the Galaxy, as well as the Small and Large Magellanic Clouds. The dashed, red line shows a linear fit to this data (Igoshev et al., 2021), and is given by the equation

$$\log_{10}\left(\frac{P_{\rm spin}}{\rm s}\right) = 1.12 \times \log_{10}\left(\frac{P_{\rm orbital}}{\rm d}\right) - 0.298. \tag{1.1}$$

Figure 1.4 shows a schematic of a BeXRB; a NS in a fairly close, moderately eccentric orbit around a Be star, which feeds a semi-stable, circumstellar decretion disc. The NS disrupts the disc during the periastron passage, causing an X-ray flare which can last for several days. The NS spins asynchronously, causing regular X-ray pulsations at the spin period of the NS.

The exact mechanism of mass loss from the Be star is still not known. Be stars are rapidly rotating, at $\gtrsim$75% of the critical rate, and a possible mechanism which can eject material from the equatorial region into a Keplerian orbit involves non-radial pulsations and small-scale magnetic fields, such as localised loops (see Rivinius et al., 2013, for a review about Be stars). An alternative mechanism involving magnetic reconnection and no non-radial pulsations has also been proposed by Balona and Ozuyar (2021)

Optical emission from BeXRBs is dominated by the companion and its decretion disc, while the accretion disc around the NS makes a contribution that is smaller by orders of magnitude. Therefore, photometric variability mainly contains information about the Be star behaviour. Optical light curves show long-term aperiodic variability caused by the evolving decretion disc,

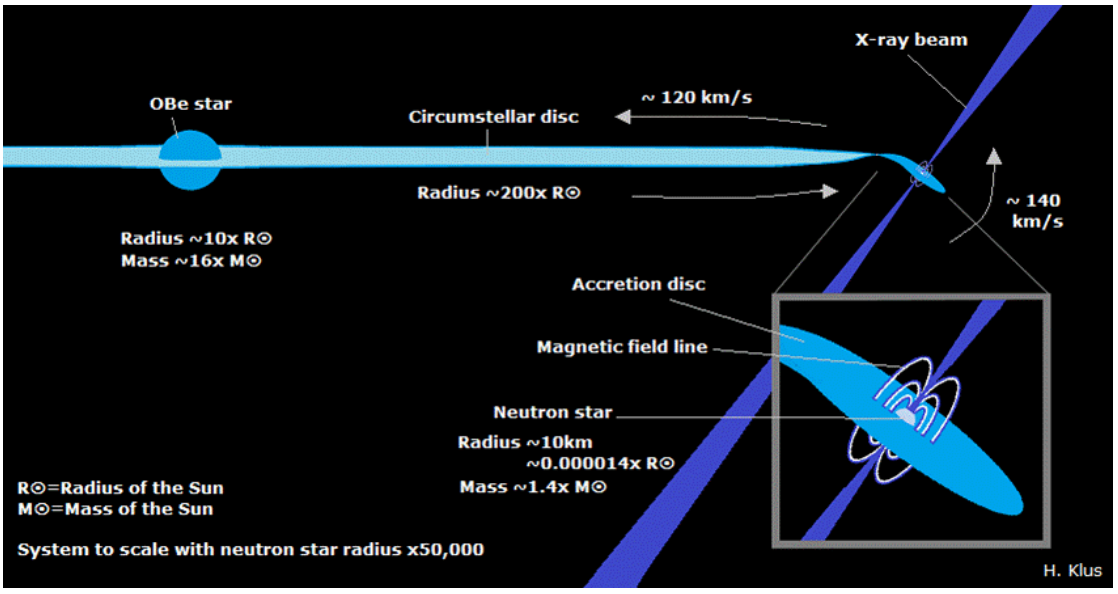


FIGURE 1.4: Be X-ray binary. The schematic depicts a system with eccentricity of 0.3 and orbital period of 150 days. Credit: Helen Klus (http://www.soton.ac.uk/~mjcoe/smc/)

periodic flares with a fast rise and exponential decay profile caused by the NS disrupting the decretion disc, as well as faster modulations with periods in the range of 0.1-2 days. Those fastest periods are associated with the pulsation and rotation of the Be star (Rivinius et al., 2013).

Chapter 4 of this thesis focuses on the optical variability of a sample of Small Magellanic Cloud BeXRBs in the timescales of days to years.

1.2.3 X-ray binaries: formation and evolution

XRBs form from binary star systems, when the more massive of the two stars undergoes a supernova explosion, where the envelope of the star is expelled, and the core of the star collapses to form a compact object. In cases where more than a half of the binary mass is expelled, the eccentricity of the binary orbit increases $e > 1$, the compact object becomes unbound. Since the more massive star explodes first, the binary systems are most likely to become disrupted, unless mass transfer between the stars takes place prior to the supernova. Formation scenarios differ for LMXBs and HMXBs (Verbunt, 1993).

The companion stars in an LMXB has a low mass of $\leq 1 M_{\odot}$, and winds are not an effective enough mechanism to bring the mass of the massive star below that value. It has been proposed that this could happen through the spiral-in mechanism, and a correctly aimed kick velocity, caused by the asymmetry of the collapse of the core of the star. Alternative scenarios include the formation of a neutron star from a white dwarf, which can implode due to the accretion-induced collapse (Whelan, 1973; Verbunt, 1993). The final scenario is the evolution of the binary from a system of multiple stars (Verbunt, 1993).

The spiral-in process involves the formation of a common envelope for the binary system. This can happen when the massive star expands rapidly, or when the distance between the stars decreases due to mass transfer. The smaller star then becomes engulfed by the massive star, and spirals in, due to the drag force causing a dissipation of the orbital angular momentum into the envelope. The envelope starts to rotate and heats up, and often gets ejected, when the orbital energy released exceeds the binding energy of the envelope (Lewin and van den Heuvel, 1983; Verbunt, 1993). After that, the close binary star will collapse and explode in a supernova, if it is massive enough.

Similarly to the LMXBs, the formation of HMXBs also requires mass transfer in order to keep the binary intact. However, the formation of HMXBs requires binary stars of high mass, so the mass transfer does not need to be as extreme. In these systems, the mass is transferred from the more massive star to the companion through Roche-lobe overflow. The more massive star is stripped, leaving a naked helium star, which then undergoes the supernova explosion (Tauris and Van Den Heuvel, 2003).

1.2.4 The origin of X-ray emission

This thesis focuses on the temporal variability of XRBs in particular energy bands, and as such it contains little consideration of the energy spectral characteristics of discussed sources, but it is worth reviewing the physical processes which produce the emissions that we observe.

Gas can escape the companion star and fall onto the compact object in the process of accretion. There are two main mechanism of accretion, and they are usually associated with the type of the system; massive companion stars in HMXBs can be more unstable and eject gas as stellar wind, and some of this wind can be captured by the compact object. In case of LMXBs, the main mechanism of accretion involves the Roche lobe overflow, where matter leaves the companion star through the inner Lagrangian point, as a result of the star filling its Roche lobe, due to either growing in size or the orbital radius shrinking.

This gas falls towards the compact object, following the steep gradient of the gravitational field, and releasing gravitational potential energy, mainly in the form of electromagnetic radiation. Energy released is given by

$$E_{\mathrm{acc}} = \frac{GMm}{R}, \tag{1.2}$$

where G is the gravitational constant, M and R are the mass and radius of the compact object, and m is the mass of the accreted gas. Electromagnetic power (luminosity) generated by accretion depends on the mass accretion rate, and hence is given by the time derivative of E_{acc}. However, not all the gravitational potential energy of accreted matter is released as radiation, and the radiative efficiency of this process can be accounted for with the parameter η;

$$L_{\mathrm{acc}} = \frac{GM\dot{M}\eta}{R}. \tag{1.3}$$

where $\dot{M} = dm/dt$. The value of η depends on the density of the compact object, and it is usually assumed to be $\eta \approx 0.1$ for NSs and stellar mass BHs (Frank et al., 2002). Accretion rate is modulated by the radiation pressure of emitted photons, leading to a hypothetical accretion rate limit which is called the Eddington limit, in case of the assumed spherically symmetric accretion of fully ionised hydrogen. Luminosity reached at the Eddington limit is given by

$$L_{\mathrm{Edd}} = \frac{4\pi G M m_{\mathrm{p}} c}{\sigma_{\mathrm{T}}}, \tag{1.4}$$

where m_{p} is proton mass, c is the speed of light, and σ_{T} is the Thomson scattering cross-section of free electrons.

However, accretion onto compact objects generally is not spherically symmetric. Due to the conservation of angular momentum, matter ejected from the companion does not fall onto the compact object directly, but rather follows a circular orbit with radius R_c dictated by its specific angular momentum.Viscous torques of gas particles cause them to transfer the angular momentum and spiral inwards, while the gravitational potential energy is dissipated in the form of heat and electromagnetic radiation.

Accreting matter forms a disc, and disc annuli at increasing radii emit thermal radiation of decreasing energy, resulting in a superposition of black body spectral shapes which peaks at ~1 keV. XRBs also emit non-thermal X-rays with higher energies of up to ~150 keV, which produce a power-law-shaped spectral component. Power-law X-rays are thought to be generated by inverse Compton scattering of thermal disc photons by hot plasma electrons, which takes place in the corona, a region close to the compact object. Photons scattered in the corona illuminate the disc and produce the reflected X-ray component, whose most prominent features are the K-alpha iron emission line at 6.4 keV, and the so-called Compton hump at 10-30 keV produced by down-scattering of power-law photons (see Gilfanov, 2010, for a review of X-ray emission from BH XRBs).

In general, the analysis of emission from XRBs can tell us about the geometry of the binary system and the accretion flow, about the physics of matter in extreme gravitational fields, and the dense matter of NSs, their equations of state. These questions can be probed with an array of different methods from the realms of energy-spectral and timing analyses. In this thesis, I focus on the latter component, and look at methodologies that can be applied to the study of XRB emission variability in time. In the next section, I provide a brief summary of methods commonly used in astronomic timing analysis.

1.3 Periodicity characterisation

The most fundamental tool used for the analysis of signals in frequency domain is the power spectrum; the modulus of the Fourier transform of a time series. However, calculation of a power spectrum requires sampling that is evenly spaced in time. In case of unevenly sampled

time series, a common method for the identification of periodic signals is the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982).

1.3.1 Lomb-Scargle Periodogram

Lomb-Scargle periodogram is a particularly popular tool in the field of astronomy, where time series are virtually always sampled unevenly and have gaps in them due to observational constrains. In this work, I use the Lomb-Scargle periodogram implementation from a community-developed Python package Astropy (The Astropy Collaboration et al., 2022), which is based on the code presented in Vanderplas et al. (2012); Vanderplas and Ivezić (2015). Specifically, I use the implementation of the generalised Lomb-Scargle (also known as floating-mean) method introduced by Zechmeister and Kürster (2009). When performing standard Lomb-Scargle analysis, it is assumed that the data is centred around the mean of the signal present in that data, while fitting of a generalised Lomb-Scargle periodogram does not make this assumption, so the generalised variant is recommended to account for the fact that full phase coverage of the signal cannot be guaranteed (VanderPlas, 2018).

For a mean-centred time series, comprised of pairs of time stamps and y-variable values (t_n, y_n), the standard Lomb-Scargle periodogram (Scargle, 1982), evaluated over frequencies f, is defined as:

$$P_{LS}(f) = \frac{1}{\sum_n y_n^2}\left(\frac{[\sum_n y_n \cos(2\pi f(t_n-\tau))]^2}{\sum_n y_n \cos^2(2\pi f(t_n-\tau))} + \frac{[\sum_n y_n \sin(2\pi f(t_n-\tau))]^2}{\sum_n y_n \sin^2(2\pi f(t_n-\tau))}\right), \tag{1.5}$$

where the left-most normalisation term is derived from residuals between the data and a constant model (χ_0) (Lomb, 1976), and τ is defined as

$$\tau = \frac{1}{4\pi f}\tan^{-1}\left(\frac{\sum_n \sin(4\pi f t_n)}{\sum_n \cos(4\pi f t_n)}\right). \tag{1.6}$$

τ is calculated for each test frequency, so that τ becomes $\tau + T_0$ when t_n becomes $t_n + T_0$, to ensure time-translation invariance, which is a property equivalent to that of an ordinary periodogram (i.e. the squared modulus of the discrete Fourier transform). Furthermore, in case of even sampling $\tau = 0$, the (unnormalised) Lomb-Scargle periodogram reduces to the ordinary periodogram (Scargle, 1982).

The Lomb-Scargle periodogram can be equivalently expressed in terms of a least-squares fit of a sine wave model

$$y_{\text{model}}(t; f) = A_f \sin\big(2\pi f(t-\phi_f)\big), \tag{1.7}$$

where the least-squares fit is computed with respect to amplitude A_f and phase ϕ_f for every test frequency, using the χ^2 statistic:

$$\chi^2(f) = \sum_n \left(\frac{y_n - y_{\text{model}}(t_n; f)}{\sigma_n}\right)^2, \tag{1.8}$$

where σ_n is the measurement uncertainty on y_n. The normalised periodogram can be expressed as:

$$P_{LS}(f) = \frac{1}{\chi_0^2}\left(\chi_0^2 - \min_{A_f,\phi_f} \chi^2(f)\right). \tag{1.9}$$

For the generalised Lomb-Scargle periodogram, an offset term is added to Equation 1.7:

$$y_{\text{model}}(t; f) = y_0(f) + A_f \sin\big(2\pi f(t - \phi_f)\big) \tag{1.10}$$

When calculating a Lomb-Scargle periodogram, it needs to be considered what set of model frequencies is to be tested. In case of a discrete Fourier transform of an evenly sampled time series, the set which produces the periodogram containing the maximal information about the power distribution of the sampled signal is referred to as the set of natural frequencies. They are defined as

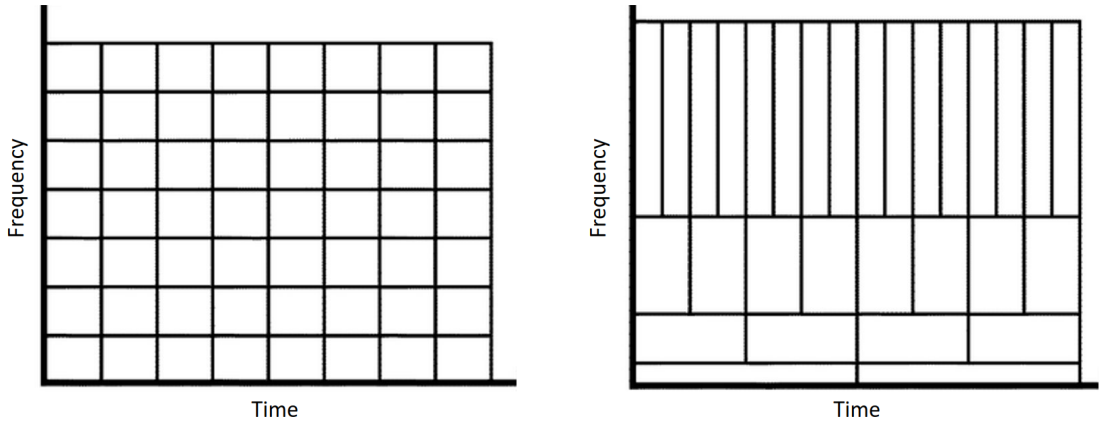
$$f_k = \frac{k}{T} \quad \text{for} \quad \{k = 1, \ldots, \frac{N_0}{2}\}, \tag{1.11}$$

where N_0 is the number of observations, and T is the time span covered by the data. In case of even sampling, this set spans frequencies between the fundamental frequency $f_1 = 1/T$, and the Nyquist frequency, $f_N = \frac{1}{2}N_0/T$ which is equal to half the sampling frequency.

In case of uneven sampling, the highest frequency limit is not well-defined, and in principle it is possible to test for frequencies higher than the pseudo-Nyquist frequency (which in this case could be generalised to some type of average of time gaps between consecutive observations), up to the frequency given by the minimal time gap between observations $f_{\text{max}} = \frac{1}{2}\Delta t_{\text{min}}$ (Scargle, 1982), however it becomes difficult to interpret the result for those high frequencies due to the effect of the convolution of the spectral window function with the spectrum of the signal, which causes spurious peaks.

Another consideration is the resolution of the frequency grid; in a time series with a baseline of T we would expect periodogram peaks of width $\sim 1/T$, so a periodogram with frequency steps of $1/T$ would not be enough to resolve them. Therefore, it is a common approach to oversample the periodogram by increasing the number of frequency steps by some factor between 5-10 (VanderPlas, 2018). Increasing the oversampling rate helps to extract more information about the power distribution, up to the limit dictated by the finite resolution of signal sampling, but it also increases the computational cost. Additionally, increasing the number of tested frequencies increases the chance of false detections due to the multiple comparisons problem.

Periodograms of poorly sampled and/or unevenly sampled, noisy time series, often show multiple spurious peaks, so it is important to be able to assess the significance of any detection. This can be done through the false-alarm probability (FAP) calculation, which is defined as the probability of generating a peak of a given magnitude in a periodogram of a data set containing only noise and no periodic signal. Analytic solutions for FAP exist, e.g. Baluev (2008) derived a formulation



(a) Short time Fourier transform (b) Wavelet transform

FIGURE 1.5: Time-frequency planes illustrating the resolution dependence in (1.5a) short time Fourier transform and (1.5b) wavelet transform.

for the upper limit of FAP which is valid for alias-free periodograms. In case of strong aliasing, it is best to apply a computational method like bootstrapping, which takes into account the effect of the sampling pattern.

The bootstrapping method for FAP calculation involves random re-sampling with replacement of flux values from the set of observed values, while keeping the temporal structure of the light curve the same. The distribution of maximum power produced by periodograms of many light curves shuffled in this way approximates the distribution of power for a light curve with the same observing window function but no periodic signal. Therefore, when estimating FAP with the bootstrapping method, the null hypothesis that we are testing against is that the peak observed in the periodogram was generated by Gaussian white noise. FAP can be used to judge the probability of a spurious peak being created by random noise, and hence inform the confidence of signal detections.

1.3.2 Weighted wavelet Z-transform

The periodogram is a useful tool for the characterisation of signals present in the data. However, in order to calculate it, we transform the data to frequency space and hence lose all the information about the location of those signals in time. For non-stationary signals, we have no way of knowing whether a signal is persistent throughout the time series or appears in certain intervals, what the order of appearance of signals might be, or in case of frequency shifts, whether the rate of frequency change was positive or negative. In order to address such questions, we need to retain some time and frequency information about the data at the same time.

A classic method that can be applied to evenly sampled data is the short time Fourier transform (STFT, Gabor, 1946), which breaks down the time series into segments and Fourier transforms them individually. This way, we can narrow down the location of the signal to a particular segment. However, this approach forces us to make a compromise between the time resolution

and the low frequency limit; shorter segments give a better time resolution but impose a more restrictive high-pass frequency filter (and also restrict the frequency resolution according to the Uncertainty Principle).

The wavelet transform uses a series of briefly oscillating functions which become shorter as their frequency increases. Figure 1.5a shows the time-frequency plane of a STFT, and a qualitative representation of the resolution grid, which is constant for all frequencies and time windows. In contrast to STFT, both time and frequency resolutions of the wavelet transform are variable; low frequency wavelets produce a high frequency resolution and a coarse time resolution, and conversely, high frequency wavelets give low frequency resolution and a high time resolution. The wavelet transform is more suitable for the time-frequency analysis of data which can contain a wide range of signal frequencies.

The STFT and wavelet transform are both performed on evenly sampled time series. The Lomb-Scargle periodogram can be used to create a dynamic periodogram as an alternative to STFT for unevenly sampled data (for examples of use, see Clarkson et al., 2003; Parkinson et al., 2003). The weighted wavelet Z-transform (WWZ) is a wavelet transform alternative developed by Foster (1996) for unevenly sampled data. To describe the WWZ of a time series $x(t)$, Foster (1996) frames the wavelet transform as a weighted projection onto a set of three trial functions:

$$\phi_1(t) = \mathbf{1}(t), \tag{1.12}$$

$$\phi_2(t) = \cos\left(\omega(t-\tau)\right), \tag{1.13}$$

$$\phi_3(t) = \sin\left(\omega(t-\tau)\right), \tag{1.14}$$

which is performed for every combination of frequency ω and wavelet time shift τ (which, for example, can be evaluated at evenly spaced locations across the time series or at every data point), and where $\mathbf{1}(t) = 1$ for all t. Projecting the data onto the trial functions involves constructing a model function $y(t)$ from the sum of products of the trial functions and model coefficients, such that

$$y(t) = \sum_{a=1}^{3} y_a \phi_a(t), \tag{1.15}$$

where the model coefficients are found by an ordinary least-squares fit to the data

$$y_a = \sum_{b=1}^{3} S_{ab}^{-1} \left\langle \phi_b | x \right\rangle, \tag{1.16}$$

where S^{-1} is the inverse of the S matrix, which is the inner product of the trial functions

$$S_{ab} = \left\langle \phi_a | \phi_b \right\rangle, \tag{1.17}$$

and the bra-ket notation $\langle f|g\rangle$ represents the weighted inner product of two arbitrary functions $f(t)$ and $g(t)$ as defined by Foster (1996):

$$\langle f|g\rangle = \frac{\sum_{\alpha=1}^{N} w_\alpha f(t_\alpha) g(t_\alpha)}{\sum_{\beta=1}^{N} w_\beta}, \tag{1.18}$$

where statistical weights are assigned to data according to

$$w_\alpha = \exp\Big(-c\omega^2(t_\alpha - \tau)^2\Big), \tag{1.19}$$

where c is the decay rate of the wavelet, which can be treated as a parameter or as a constant; a conventional value for c when dealing with light curves of variable stars is $1/8\pi^2 \approx 0.0125$ (Foster, 1996). The statistical weights reflect the amplitude of the abbreviated Morlet wavelet at a time t. The abbreviated Morlet wavelet is the analysing wavelet used in the WWZ, and following the notation convention from Foster (1996), it is defined as

$$f(z) = \exp\Big(iz - cz^2\Big) = \exp\Big(i\omega(t-\tau) - c\omega^2(t-\tau)^2\Big). \tag{1.20}$$

The wavelet peaks at $z = 0$, and therefore is the most sensitive to the behaviour of data close to $t = \tau$, and decays with a Gaussian profile as we move away from τ.

The WWZ power is then defined as

$$Z = \frac{(N_{\text{eff}} - 3)V_y}{2(V_x - V_y)}, \tag{1.21}$$

where N_{eff} is the effective number of data points, which accounts for the changing number of sampled data points as the position and frequency of the wavelet changes, and is defined as

$$N_{\text{eff}} = \frac{(\sum_\alpha w_\alpha)^2}{\sum_\alpha w_\alpha^2}, \tag{1.22}$$

V_x is the weighted variation of the data

$$V_x = \frac{\sum_\alpha w_\alpha x^2(t_\alpha)}{\sum_\lambda w_\lambda} - \left(\frac{\sum_\alpha w_\alpha x(t_\alpha)}{\sum_\lambda w_\lambda}\right)^2 = \langle x|x\rangle - \langle \mathbf{1}|x\rangle^2, \tag{1.23}$$

and V_y is weighted variation of the model function

$$V_y = \frac{\sum_\alpha w_\alpha y^2(t_\alpha)}{\sum_\lambda w_\lambda} - \left(\frac{\sum_\alpha w_\alpha y(t_\alpha)}{\sum_\lambda w_\lambda}\right)^2 = \langle y|y\rangle - \langle \mathbf{1}|y\rangle^2. \tag{1.24}$$

Foster (1996) points out that WWZ gives a good measure of signal frequency and a poor measure of amplitude, and so provides an additional metric of the weighted wavelet amplitude (WWA), the square root of the fitted sine and cosine coefficients:

$$WWA = \sqrt{(y_2)^2 + (y_3)^2}, \tag{1.25}$$

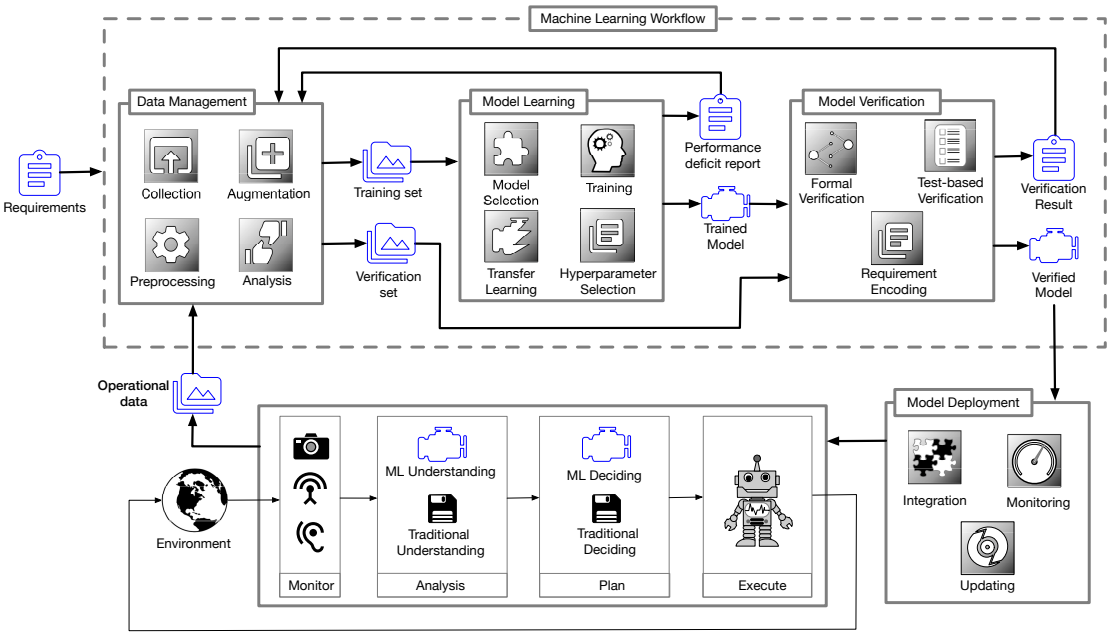


FIGURE 1.6: A generalised lifecycle of an ML solution. The major components include; the Monitor-Analyse-Plan-Execute loop (bottom-left), the Machine Learning Workflow (top), and the Model Deployment phase (bottom-right). Credit: Ashmore et al. (2021)

which can be used to find the amplitude of a signal after applying WWZ to find the frequency of that signal.

1.4 Machine learning methods

Chapters 2 and 3 of this thesis focus on the development and application of ML tools for the automated feature extraction and classification of light curves. In this section, I introduce the fundamental ML concepts and algorithms which constitute parts of the pipeline which can extract insights from astronomical time series data.

1.4.1 General machine learning workflow

As discussed at the beginning of this chapter, ML solutions are often deployed to tackle problems within complex, dynamic systems, and aim to replace traditional data analysis and decision-making with autonomic elements, capable of producing actionable knowledge without human supervision.

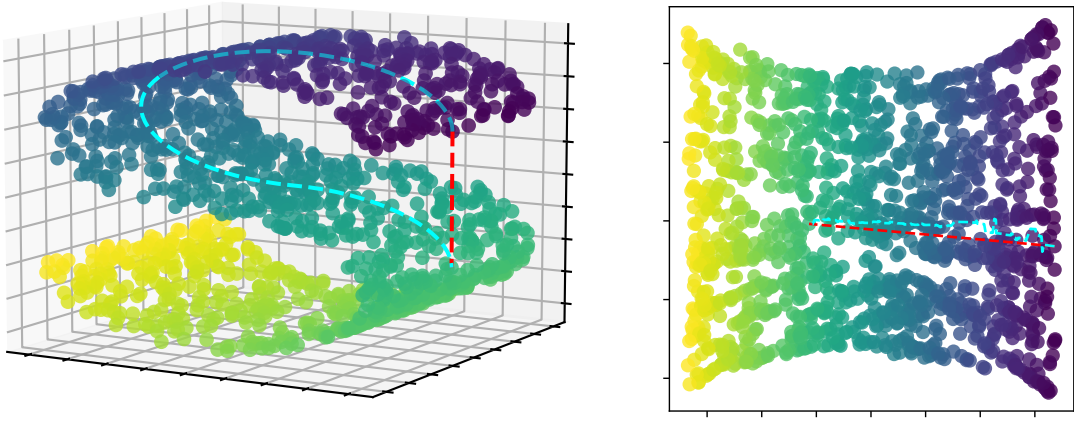
In general, the development process of an ML model consists of a few fundamental phases, including the data acquisition and preprocessing phase, model training and selection phase, followed by the model testing phase. The flowchart in Figure 1.6 shows this process within the Machine Learning Workflow (top block). However, the lifecycle of an ML solution does not end there; the fourth stage of the lifecycle is the Model Deployment (bottom-right block of

Figure 1.6), where the solution is implemented to tackle the problem that it was designed for. The deployed ML solutions can replace the human in multiple stages of the Monitor-Analyse-Plan-Execute (MAPE) loop In a MAPE loop monitoring systems (i.e. sensors) are used to gather data from a system of interest. Data then needs to be encoded into a machine-readable representation, comprehended and analysed. Once actionable information is extracted from the data, it can be used by autonomous systems for decision-making and execution of adequate actions. Operation of the MAPE loop by traditional operators or autonomous agents leads to the production of data and the formulation of requirements for a new or improved autonomous agent, which can be leveraged in the next iteration of development, in the Machine Learning Workflow (for more details about the MAPE loop, for example see Kephart and Chess, 2003). An example system that could use ML to perform autonomous actions is an automatic source classification system for an astronomical survey, which alerts the users when an interesting type of source is detected.

1.4.2 Representation learning

It is rarely the case for data to be used successfully in ML applications while in the form that it was originally recorded in. One aspect which is crucial for the successful application of ML is the construction of data representations which meet few important criteria; they must be compatible with the algorithm to be used, and informative enough so that the ML task could provide accurate results based on the available data sample. This is why the "ML Understanding" stage is an integral part of the lifecycle of an ML solution, shown in Figure 1.6. Aligning the data representation with these requirements is usually achieved in the process of feature engineering. In this process, the analyst aims to construct standard feature vectors, where each observation in the data set is represented by a constant number of values; only then it becomes possible to compare observations with each other and make inferences using ML. The number of components of the feature vector effectively defines the dimensionality of the feature space occupied by the data observations. The arrangement of data in this space reflects how informative the representation is; random arrangement of data in the feature space offers no information and does not allow for any inference to be made about any particular observation. In the case of an informative feature space, data is organised in a way that allows the analyst to discriminate observations of different categories and/or quantify the similarity of observations based on the distance between them. This organisation of data in the feature space can manifest in the formation of clusters or manifolds.

A manifold is a topological space that is locally Euclidean. Manifold hypothesis poses that high-dimensional data tend to lie along low-dimensional manifolds embedded in the high-dimensional space (Fefferman et al., 2016), implying that it is possible to represent the high dimensional data with a lower-dimensional feature vector. Figure 1.7 shows how three-dimensional data of a two-dimensional manifold can be reduced to two variables, which capture all the relevant variability in that data. This visualisation also illustrates how Euclidean distance in the original feature space of the data might not be an adequate metric for describing dissimilarity between observations



(a) Original data (b) Isomap Embedding

FIGURE 1.7: Two-dimensional data manifold embedded in a three-dimensional feature space. Figure 1.7a shows the original S-shaped data cloud in three dimensions, while Figure 1.7b shows the output of a manifold learning algorithm Isomap. Axes of all panels use arbitrary units. Dashed red lines show the Euclidean distance between two arbitrary data points in the 3D space, and the cyan lines show the distance between the same pair of points along the manifold surface. The embedding is generated using SCIKIT-LEARN implementations of the algorithm (Pedregosa et al., 2011).

over long distances; the Euclidean distance between two arbitrary data points at the right edge of the data cloud is indicated by the dashed red line, and due to the curvature of the manifold, it is not a good metric of similarity between those data points. A more adequate distance metric would be the geodesic distance indicated by the dashed cyan curve, which reflects the true geometry of the low-dimensional manifold.

An appropriate dimensionality reduction algorithm can help to approximate the distance along the manifold surface. Figure 1.7b shows the data from Figure 1.7a embedded in two-dimensional space using the Isometric Mapping embedding (Isomap, Tenenbaum et al., 2000). Isomap embedding shows the Euclidean and geodesic paths joining the same two data points as Figure 1.7a. The Isomap embedding is effective at flattening continuous manifolds, and thanks to the embedding in the low-dimensional space, Euclidean distance becomes a much better approximation of the true geodesic distance. The geodesic distance across the Isomap projection is represented as a series of lines instead of a smooth curve because it traces the edges of the graph generated by the algorithm, which connects neighbouring data points. However, Isomap might not be as efficient at separating multiple manifolds into individual clusters, and such behaviour might be observed in embeddings created with algorithms like the t-distributed Stochastic Neighbour Embedding (t-SNE, Van Der Maaten and Hinton, 2008).

The goal of dimensionality reduction is to identify the minimal number of parameters which describe the position on this true low-dimensional manifold, and capture the intrinsic degrees of freedom of the data, effectively removing the random, uninformative variability from the feature set. This approach assumes that the high-dimensional data is generated by a process controlled by a small number of parameters, referred to as latent variables. In this work, I focus on the

development of an autoencoder; a method which extracts the latent variables using a neural network. More details about neural networks, and their application to dimensionality reduction, can be found in Section 1.4.3.1 and chapters 2 and 3.

1.4.2.1 Uniform Manifold Approximation and Projection

Uniform Manifold Approximation and Projection (UMAP, McInnes et al., 2018) is a manifold learning and dimensionality reduction algorithm used extensively in this thesis for the visualisation of high-dimensional data, so it requires a brief introduction. UMAP constructs the low-dimensional embedding in two steps; in step (1) it builds a fuzzy topological representation of the input data set, and then in step (2) it finds a low dimensional embedding which minimises the differences between the low-dimensional fuzzy topological representation and the high-dimensional counterpart.

Step (1): As McInnes et al. (2018) frame this, the topological representation can be constructed using a simplicial complex, i.e. a combination of simplexes. Individual data points are 0-simplexes, pairs of connected points trace out 1-simplexes, three connected points trace out 2-simplexes and so on (k-simplex is the convex hull of $k+1$ points). In order to construct higher dimensional simplexes, we need to set a distance limit below which data points are allowed to connect.

However, UMAP assumes that the input data sample is distributed uniformly across the manifold, which is rarely the case in practical applications. Therefore, the algorithm equivalently assumes that the space between neighbouring data points expands or shrinks, depending on the distance between those data points. This makes the connectivity problem equivalent to building a nearest neighbour graph. The graph is constructed using the K-nearest neighbour graph algorithm by Dong et al. (2011). The hyperparameter *n_neighbors* controls the number of neighbours found for each data point, whereas the hyperparameter *local_connectivity* sets the minimal number of neighbours which are required to be connected to each point. The edges of the graph are weighted according to the probability of a connection between two data points existing, and the value ranges between 0 and 1, where the number of nearest neighbours dictated by the *local_connectivity* hyperparameter always has the value of 1, and the data points further than the number set by the *n_neighbors* hyperparameter always have the value of 0, whilst the rate of decay between those two extremes depends on the distance function of a given data point. Since the distance function of a data point depends on the distance to the k-th nearest neighbour, the distance from point a to point b can be different from the distance from b to a, resulting in two edge weights, w_{ab} and w_{ba}. Therefore, weights are combined using the following formula;

$$w = w_{ab} + w_{ba} - w_{ab} \cdot w_{ba}. \tag{1.26}$$

This probabilistic approach to graph construction results in the fuzzy logic in the topological representation of the data.

Step (2) of the UMAP is to embed this fuzzy topological structure in a low-dimensional, (usually) Euclidean space, find the fuzzy topological structure in the embedded space, and minimise the difference between the two. This is done through minimisation of the cross-entropy cost function C_{UMAP}, which compares the weights of graph edges of the high and low dimensional topological representations;

$$C_{\mathrm{UMAP}} = \sum_{e \in E} w_h(e) \log\left(\frac{w_h(e)}{w_l(e)}\right) + (1 - w_h(e)) \log\left(\frac{1 - w_h(e)}{1 - w_l(e)}\right), \tag{1.27}$$

where $w_h(e)$ and $w_l(e)$ are the weights of edge e in high and low dimensional representations respectively, where e is one from the set of all edges E. The cross-entropy cost function is chosen because the sets of edge weights describe the probabilities of 1-simplexes existing, and as such they follow the Bernoulli distribution. This choice of the cost function also frames the algorithm for the construction of the low-dimensional fuzzy topological structure as a force directed graph layout algorithm. The first term of Equation (1.27) is an attractive force, because it is minimised when the value of $w_l(e)$ is maximised, which happens when data points which constitute the 1-simplex e are positioned close together in the low-dimensional representation. The second term of Equation (1.27) is a repulsive force, minimised when the value of $w_l(e)$ is minimised, which happens when the data points are set far apart. The balance of these two forces rearranges the low-dimensional representation, which should settle in a state that is a relatively accurate representation of the topology of the original data. The cost function C_{UMAP} is minimised using the stochastic gradient descent (for more details about stochastic gradient descent, see for example sec. 5.9 of Goodfellow et al., 2016)

This switch from relative to absolute distance metric in step (2) of UMAP requires the adoption of another hyperparameter *min_dist*, which controls the minimal distance between data points and hence regulates the density of data in the projection. The dimensionality of the final embedding is controlled by the *n_components* hyperparameter.

The hyperparameters of UMAP visualisations presented in this thesis follow the default values recommended by McInnes et al. (2018); *n_neighbors* = 15, *local_connectivity* = 1, *n_components* = 2, *min_dist* = 0.1.

1.4.3 Artificial neural networks

In this thesis, UMAP algorithm is only used as a dimensionality reduction tool for visualisation purposes. Chapters 2 and 3 focus on the application of a specific type of dimensionality reduction algorithm for the purpose of feature extraction for quantitative data analysis. The algorithm is called Long-Short Term Memory Variational Autoencoder (LSTM-VAE), and it combines a couple of type of artificial neural networks. In this section, I introduce concepts required to understand how the LSTM-VAE algorithm works.

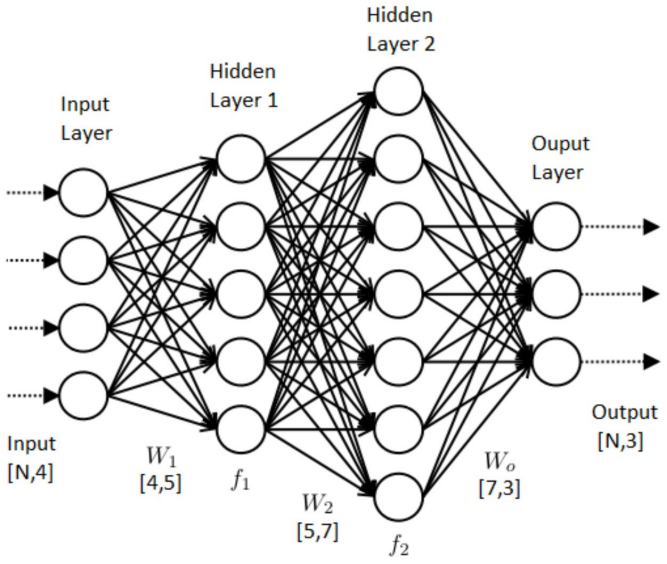


FIGURE 1.8: An example of a fully connected feedforward neural network architecture. N is the batch size of the input. Credit: Baron (2019)

Artificial neural networks (ANNs) are a family of machine learning algorithms inspired by the structure of biological neurons (see McCulloch and Pitts, 1943, for one of the first works on the topic). ANN can be represented as directed, weighted graphs, with the nodes arranged in layers. Nodes represent neurons, which perform computation and feed their outputs to the nodes in the following layers, whilst the weights on edges represent the strength of connections between the neurons. ANNs are very flexible in terms of structure and can be applied to a variety of tasks, including classification, regression, dimensionality reduction and clustering. Figure 1.8 shows an exampled of a basic ANN architecture, which, in the order of data flow, consists of an input layer, two hidden layers and an output layer. Every neuron beyond the input layer sums the weighted output of the neurons from the previous layer, adds a bias term, and applies a non-linear activation function. Therefore, if $\vec{x}$ is the input vector, then the output of the first hidden layer (i.e. the hidden state) is given by

$$\vec{h_1} = f_1(W_1\vec{x} + \vec{b_1}), \tag{1.28}$$

where f_1 is a non-linear activation function of the first hidden layer, W_1 is the weight matrix for connections between the input and first hidden layers, and $\vec{b_1}$ is the vector of bias values of the first hidden layer. Analogically, the output of the second hidden layer is computed from the $\vec{h_1}$ vector, and corresponding weights and biases. Some popular activation functions include the sigmoid function

$$\sigma(a) = \frac{1}{1 + e^{-a}}, \tag{1.29}$$

hyperbolic tangent function

$$\tanh(a) = \frac{e^{2a} - 1}{e^{2a} + 1}, \tag{1.30}$$

and rectified linear unit function

$$\mathrm{ReLU}(a) = \max(a, 0). \tag{1.31}$$

In spite of this simple structure of the ANNs, their flexibility and non-linearity allows them to approximate extremely complex mappings between the input and output. In fact, they are general function approximators, capable of approximating any functional mapping to arbitrary precision, given a large enough number of hidden neurons, even when arranged into a single hidden layer, a so-called "shallow" neural network (Cybenko, 1989; Hornik, 1991). However, deep networks with multiple layers tend to be more efficient at function approximation, in terms of the number of parameters required (Eldan and Shamir, 2016).

Weights and biases are free parameters of the ANN, whilst the number and arrangement of neurons as well as the choice of the activation functions and the cost function are hyperparameters selected prior to model fitting. The process of finding the parameter values which result in the best function approximation is usually referred to as model training, and this is commonly achieved using the backpropagation and stochastic gradient descent (SGD) algorithms.

Training of a neural network is an iterative process, where the training data set is propagated through the network in batches. The output of the network is compared to the target output using a cost function (e.g. mean square error), and the backpropagation algorithm is used to calculate the gradient of the loss function with respect to the ANN parameters, working its way backwards through the network. SGD is then used to adjust the parameters in accordance with the calculated gradients. SGD is an extension of the gradient descent algorithm, which adjusts parameter values following the minimal first order derivative of the cost function with respect to the parameters. A stochastic variant of the algorithm is used because of the prohibitive computational cost of calculating the gradient for every observation in the training data set. Training of ANNs, especially deep ANNs, requires large amounts of data, and SGD capitalises on the fact that the calculated gradient is an expectation value that can be estimated from a small sample of the data. Therefore, parameter updates happen only once per training batch, using mean gradients estimated from a batch, where the batch is randomly sampled from the training data set without replacement. A training epoch finishes once all the training data is passed through this process, and multiple epochs are usually required to minimise the loss function. For more details about ANNs, see Goodfellow et al. (2016).

1.4.3.1 Autoencoders

As we have established earlier, the ultimate goal of feature engineering is to construct the minimal set of features which capture the variability of data. It often requires domain specific expertise from the analyst, who needs to identify descriptors containing relevant information about the observations in the data set (for example see Richards et al., 2011). Automated feature extraction is an alternative to manual feature engineering, and requires less specific domain knowledge. Automated feature extraction often involves methods which encode data into a more abstract, low-dimensional, latent representation. The resulting latent representation contains compressed information about the input data, and reflects the similarities and differences

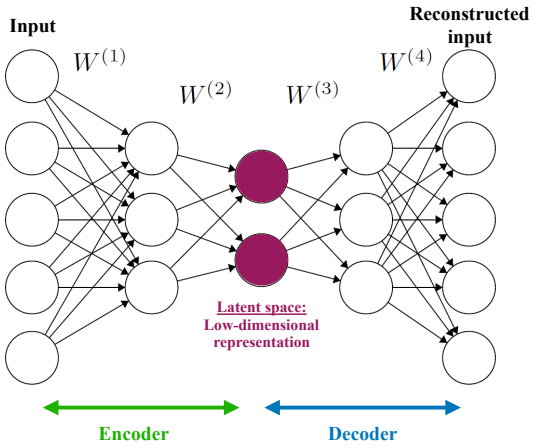


FIGURE 1.9: An example of an autoencoder architecture. Credit: Baron (2019)

between observations. ML algorithms can leverage this information to perform a variety of tasks, like classification, outlier detection or clustering.

In order to represent the light curve data using a small number of variables, and hence allow me to analyse that data space in a relatively small number of dimensions, I perform dimensionality reduction using a neural network. This process aims to extract a small set of latent or "hidden" variables, which encode information about the structured variability of the light curves. Figure 1.9 shows the basic structure of an autoencoder ANN; it consists of two blocks, an encoder and a decoder. The encoder's task is to learn a low-dimensional representation of an input data, and encode it efficiently in the latent space. The decoder uses the latent representation to reconstruct the original input of the encoder. Therefore, the cost function of an autoencoder is a function of the input and the output of the network, and as such it does not require any additional target variable for its evaluation, which means that autoencoders can be used for unsupervised learning from unlabelled data. Due to the number of neurons decreasing between the input layer and the latent space layer, the encoder must push the maximal amount of relevant information through the bottleneck of the latent space. This forces the encoder to learn a mapping that preserves the most information about the data, and ignores any noise which is not useful for the reconstruction of the input data. This behaviour makes autoencoders a natural choice for the task of compression and dimensionality reduction. Resulting latent variables are the compressed representation of data and can be leveraged in the data analysis process. For more details about autoencoders, see Goodfellow et al. (2016).

The variational autoencoder (VAE) is a type of probabilistic neural network model first proposed by Kingma and Welling (2014). Just like the non-variational type, it uses two blocks of neurons; an encoder, which maps input data onto a small set of latent variables, and a decoder, which maps them back to the input data space, hence reconstructing observations of input data. In case of the VAE however, the latent variables are distributions (continuous latent variables), so the encoder maps input data to normal distributions instead of infinitesimal points. Latent variables

are parameterised by mean and variance values, and the decoder uses random samples from those distributions to produce input reconstructions.

VAEs use an additional term in their cost functions, the Kullback-Leibler divergence. It is a regularisation term, which prevents the distribution of the latent variables from substantially deviating from the standard normal distribution. This term is used to ensure that the latent space is continuous and meaningful. The generic Kullback-Leibler divergence between two discrete distributions $P(x)$ and $G(x)$ defined over sample space X is given by:

$$D_{\mathrm{KL}}(P \parallel G) = -\sum_{x \in X} P(x) \log \frac{P(x)}{G(x)}, \tag{1.32}$$

which in the case of two Z-dimensional Gaussian distributions can be expressed as (Kingma and Welling, 2014);

$$D_{\mathrm{KL}} = -\frac{1}{2} \sum_{j=1}^{Z} \left(\log\left(\hat{\sigma_j}^2\right) - \hat{\mu_j}^2 - \hat{\sigma_j}^2 + 1 \right), \tag{1.33}$$

for every x, where Z is the number of latent variables, $\hat{\sigma_j}^2$ is the variance of the latent variable j, and $\hat{\mu_j}$ is the mean of latent variable j.

The use of the regularisation term has several implications; deviations from the Gaussian prior are penalised, which encourages the VAE to construct less complex latent spaces. Regularisation should therefore prevent overfitting to data as the number of latent variables is increased (since sparse encoding is encouraged) (Kingma and Welling, 2014). Additionally, since the mapping to the latent space is continuous, the space tends to be smooth, and does not show abrupt changes between neighbouring locations. The VAE can generalise to previously unseen data points and interpolate across the latent space, generating new, plausible observations of data. For more details about VAEs, see Doersch (2016).

1.4.3.2 Recurrent neural networks

Recurrent neural networks (RNNs) have directed cyclic graph structure, as they build on the feedforward ANNs by adding feedback connections, which feed the output of a neuron back into itself as an additional input. This structure makes RNNs suitable for the processing of sequential input data, which is processed iteratively, one data point at a time. At every iteration, the state of the cell is fed as input of the next iteration, which allows the network to learn from consecutive points of the sequence and extract temporal patterns. The use of RNNs has been researched extensively for the processing of text, handwriting, speech and sound (Yu et al., 2019, and references therein).

Figure 1.10 shows an illustration of a basic multivariate RNN architecture. It consists of the input layer of size N (bottom), one hidden recurrent layer with M recurrent cells/neurons (middle), and an output layer of size P (top). For an input vector $\vec{x_t} = (x_1, x_2, \ldots, x_N)$, where $\vec{x_t}$ is one of

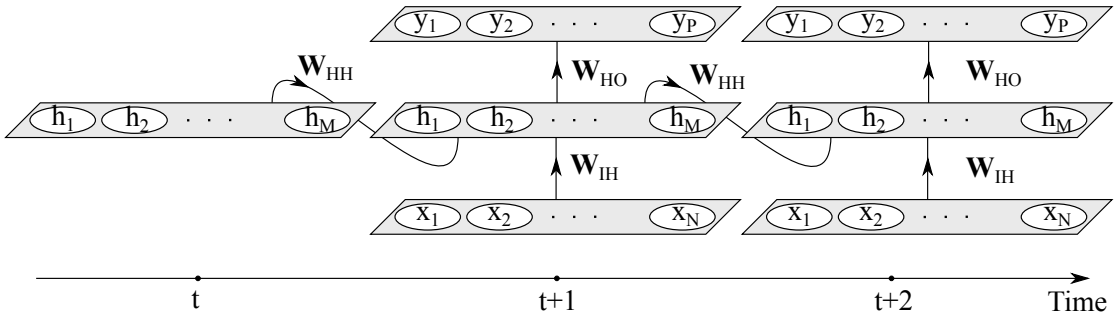


FIGURE 1.10: An example of a recurrent neural network architecture, unfolded in time. Credit: Salehinejad et al. (2017)

the series of T vectors from the input data matrix $X = (\vec{x}_1, \ldots, \vec{x}_t, \ldots, \vec{x}_T)$, so $X \in \mathbb{R}^{N \times T}$, the output of the hidden layer is given by

$$\vec{h}_t = f_H(W_{IH}\vec{x}_t + W_{HH}\vec{h}_{t-1} + \vec{b}_H), \tag{1.34}$$

where f_H is the activation function of the hidden layer, W_{IH} is the weight matrix for connections between the input and first hidden layers, W_{HH} is the weight matrix for the feedback connections between the first hidden layer and itself from the previous time step, $\vec{h}_{t-1}$ is the output of the hidden layer from the previous time step, and $\vec{b}_H$ is the vector of bias values of the hidden layer. Hidden layer state is initialised with small non-zero values at the first iteration.

RNNs build on ANN's representation of the hidden state (Equation (1.28)) by adding the self-feedback term, $W_{HH}\vec{h}_{t-1}$, which allows the hidden state of the neuron from the previous iteration to affect the output of the current iteration. Due to the cyclic nature of the RNNs, they require and alternative training algorithm, and one possibility is to apply back-propagation through time (Werbos, 1990), which unfolds the RNN and effectively treats it like an equivalent feed-forward ANN, storing all the states of intermediate iterations for the backpropagation pass.

This simple structure of RNNs allows them to model and make predictions based on the immediate context of the processed data point, but they tend to quickly forget information that is not frequently reinforced, which means that they struggle to capture long term patterns. Long-Short Term Memory (LSTM) cells, first proposed by Hochreiter and Schmidhuber (1997), address this issue through the introduction of so-called "gates" which consist of non-linear functions that control the state of the LSTM cell and protect the relevant information over long time scales.

Figure 1.11 shows a diagram of the LSTM cell, and it depicts the cell input-output flow in the vertical direction, while gates modulating the cell activity are depicted as horizontal connections. The most important component is the cell state (c_t), depicted as the black dot in the centre. It has a linear self-feedback loop, which is controlled by the forget gate. The forget gate uses the sigmoid function to set weights of the self-feedback loop to values between 0 and 1. This controls the rate at which the value stored in the cell state unit can decay. The forget gate can be defined as

$$g_t^f = \sigma(W_{g^f}\vec{x}_t + U_{g^f}\vec{h}_{t-1} + \vec{b}_{g^f}), \tag{1.35}$$

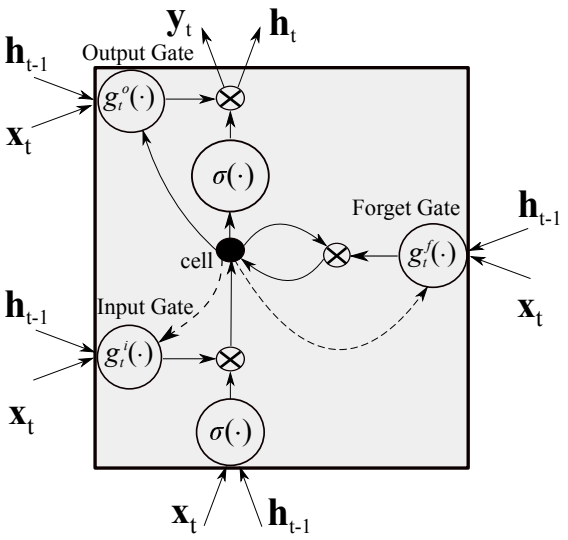


FIGURE 1.11: An illustration of a Long-Short Term Memory cell used in recurrent neural networks. Dashed lines indicate delayed (peephole) connections. Credit: Salehinejad et al. (2017)

where $\sigma(.)$ is the sigmoid function (Equation (1.29)), W_{gf} is the weight matrix from the input layer to the forget gate, U_{gf} is the recurrent weight matrix from hidden state to the forget gate, and $\vec{b}_{gf}$ is the bias vector of the forget gate. The updates to the cell state are also controlled by the input gate, which is computed similarly to the forget gate

$$g_t^i = \sigma(\mathrm{W}_{gi}\vec{x}_t + \mathrm{U}_{gi}\vec{h}_{t-1} + \vec{b}_{gi}), \tag{1.36}$$

with analogical indexing of weights and biases. The cell state c_t is then updated by the forget and input gates, following the equation

$$c_t = g_t^f c_{t-1} + g_t^i \sigma(\mathrm{W}_c\vec{x}_t + \mathrm{U}_c\vec{h}_{t-1} + \vec{b}_c), \tag{1.37}$$

where W_c is the weight matrix from the input layer to the cell, U_c is the recurrent weight matrix from the hidden state to the cell, and $\vec{b}_c$ is the bias vector of the cell state. The output of the cell is controlled by the output gate, similar to the other two gates;

$$g_t^o = \sigma(\mathrm{W}_{go}\vec{x}_t + \mathrm{U}_{go}\vec{h}_{t-1} + \vec{b}_{go}), \tag{1.38}$$

with analogical indexing of weights and biases. Finally, the output of the cell is given by

$$\vec{h}_t = g_t^o \tanh(c_t), \tag{1.39}$$

where $\tanh(.)$ is the hyperbolic tangent function (Equation (1.30)).

Thanks to the use of the forget gate, LSTMs have the potential to store information over an arbitrary time lag. The improvement of LSTMs over simple RNNs comes at the cost of increased

complexity; LSTM cells use four times as many parameters as the RNN equivalent. For more details about RNNs, see Goodfellow et al. (2016) or Salehinejad et al. (2017).

1.4.4 Other machine learning algorithms

The application of several other machine learning algorithms is discussed in following chapters of this thesis. Clustering algorithms are applied in the process of light curve feature extraction, while classification is performed on light curves based on the extracted features using the Random Forest Classifier. A brief introduction to those algorithms is provided in this section.

1.4.4.1 Gaussian mixture model

Gaussian Mixture Models (GMMs) approximate the probability distribution of the data using a weighted sum of multidimensional Gaussian components. The likelihood function of a GMM with K components is given by

$$p(\vec{x}|\theta) = \sum_{i=1}^{K} w_i g(\vec{x}|\vec{\mu}_i, \mathbf{\Sigma}_i), \tag{1.40}$$

where $\vec{x}$ is an D-dimensional data vector, w_i is the mixture weight of the ith component, and $g(\vec{x}|\vec{\mu}_i, \mathbf{\Sigma}_i)$ is the D-dimensional Gaussian density function;

$$g(\vec{x}|\vec{\mu}_i, \mathbf{\Sigma}_i) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} \exp\left\{-0.5(\vec{x} - \vec{\mu}_i)\Sigma_i^{-1}(\vec{x} - \vec{\mu}_i)\right\}, \tag{1.41}$$

where $\vec{\mu}_i$ and $\mathbf{\Sigma}_i$ are the mean (position) vector and the co-variance matrix of the i-th Gaussian component, and θ represents all the weights and parameters of the K Gaussian components.

The number of components is a hyperparameter set by the user, and the mean position of each component is initiated randomly. The position and co-variance matrix of each component are then iteratively optimised to maximise the likelihood of the data under the model using the Expectation-Maximisation algorithm (Dempster et al., 1977). GMM is a "soft" clustering method; the likelihood value of each data point is calculated for every Gaussian component, and the data points are assigned to components which give the largest likelihood. The co-variance matrices of each component can be constrained to be diagonal, tied across the components, or replaced with a single scalar value to represent spherical Gaussian components. The number of components and the type of co-variance matrix determine the complexity of the model, and they are usually adjusted based on the complexity and amount of available data. For more details about GMMs, see Reynolds (2009).

1.4.4.2 Hierarchical clustering

Hierarchical (agglomerative) clustering algorithm creates a hierarchy of clustering solutions through recursive merging of pairs of clusters. The merging algorithm starts with the value of the distance threshold set to zero, resulting in the assignment of each data point to their own singleton clusters. The threshold is then increased, and pairs of nearby clusters are merged when the distance between them is less than the threshold. The algorithm terminates when all data points belong to a single cluster, or when a threshold set by the user is reached.

The hierarchical clustering algorithm can use a variety of distance metrics and different methods of calculating distance between clusters. Clustering methods include; the "single" method, which defines the distance between two clusters as the distance between the nearest data points of the two clusters, the "average" method defines the distance between clusters as the distance between mean positions of constituent data points, and "complete" method uses the distance between two furthest data points. Another popular method is the "ward" method, which minimises the variance within clusters.

The choice of clustering method affects the resulting cluster structure, for example "single" method tends to build clusters with very uneven sizes ("rich gets richer" type of pattern), it can erroneously chain together multiple neighbouring clusters, however it performs well for non-isotropic clusters. The "average" method struggles to correctly capture elongated clusters, where one cluster can be split into multiple, or separate elongated clusters in close proximity can be erroneously merged into one. The "complete" method methods tends to create more even, compact clusters. For more details about hierarchical clustering, see Rokach and Maimon (2005).

1.4.4.3 Random forest classifier

The random forest ensemble (Breiman, 2001) is a very versatile algorithm, which can be used to perform both classification and regression, as well as unsupervised learning tasks, however here I will focus on the classification functionality. The random forest classifier uses an ensemble of decision trees, and it averages the results across the trees to find the final classification. Each decision tree is trained on a subsample of observations, picked at random from the training data set using bootstrapping. Additionally, only a subset of features is used to train each decision tree, which is another source of randomness impacting the construction of the classifier. This randomness allows the random forest to reduce over-fitting and the overall variance of the classifier, which leads to improved classification performance.

The individual decision trees of the random forest are comprised of decision nodes arranged into a branching structure, where threshold conditions are compared against the feature values of individual observations, which effectively splits the data set as it propagates through the decision tree. The terminal nodes of the decision tree are associated with data labels. During the training

of a classifier based on decision trees, the choice of the data feature and the threshold value to be used for a given decision node is made to optimise the splitting of the data set in a way that separates observations of different classes. Popular choices for the criterion which quantifies the quality of a split are Gini impurity and information gain. Gini's impurity is defined as

$$\mathrm{I}_G = 1 - \sum_{k=1}^{K} p_k^2, \tag{1.42}$$

where K is the number of classes and p_k is the proportion of observations of class k in a set. Information gain is the difference between the entropy of a set before and after a split. Entropy is defined as

$$\mathrm{E} = -\sum_{k=1}^{K} p_k log(p_k). \tag{1.43}$$

Other than the choice of the criterion for optimal data set splitting, the random forest algorithm has few relevant hyperparameters which need to be adjusted by the user. The number of decision trees in the ensemble affects the variance of the classifier; increasing the number of trees tends to reduce the variance of predictions, as the forest converges on the answer, but at the expense of increased computation time. Without any additional constrains on the decision trees, they are extended until every terminal node has only one observation in it. Therefore, the construction process of the decision trees is also controlled by setting the maximum depth of the tree, the minimal number of observations required to split a node, as well as the minimal number of observations required at a terminal node. For more information about the random forest algorithm, see Breiman (2001).

Chapter 2

Light curve fingerprints: an automated approach to the extraction of X-ray variability patterns with feature aggregation – an example application to GRS 1915+105

2.1 Introduction

The work presented in this chapter contributes to the toolbox of automated characterisation methods for light curve data. I explore the use of neural networks and Gaussian mixture models for descriptive feature generation and clustering. I construct a representation of light curve data called the "fingerprints", which can be used in downstream classification, outlier detection and clustering tasks. I demonstrate that such a method is useful in the analysis of the evolution of a particular source of interest. In this work, I analyse the complete data set of GRS 1915+105 observations captured by the Proportional Counter Array on-board of the Rossi X-ray Timing Explorer (RXTE/PCA) between 1996 and 2011.

2.1.1 Rossi X-ray Timing Explorer

Rossi X-ray Timing Explorer was a NASA satellite, launched in December 1995 and decommissioned in January 2012. It carried two pointed instruments, where the low energy range (2-60 keV) was covered by the Proportional Counter Array (PCA, Jahoda et al., 1996), comprised of five Proportional Counting Units (PCUs), with the total collecting area of 6500 cm^2. It had a collimator, a field of view with a FWHM of 1° and no angular resolution, the time

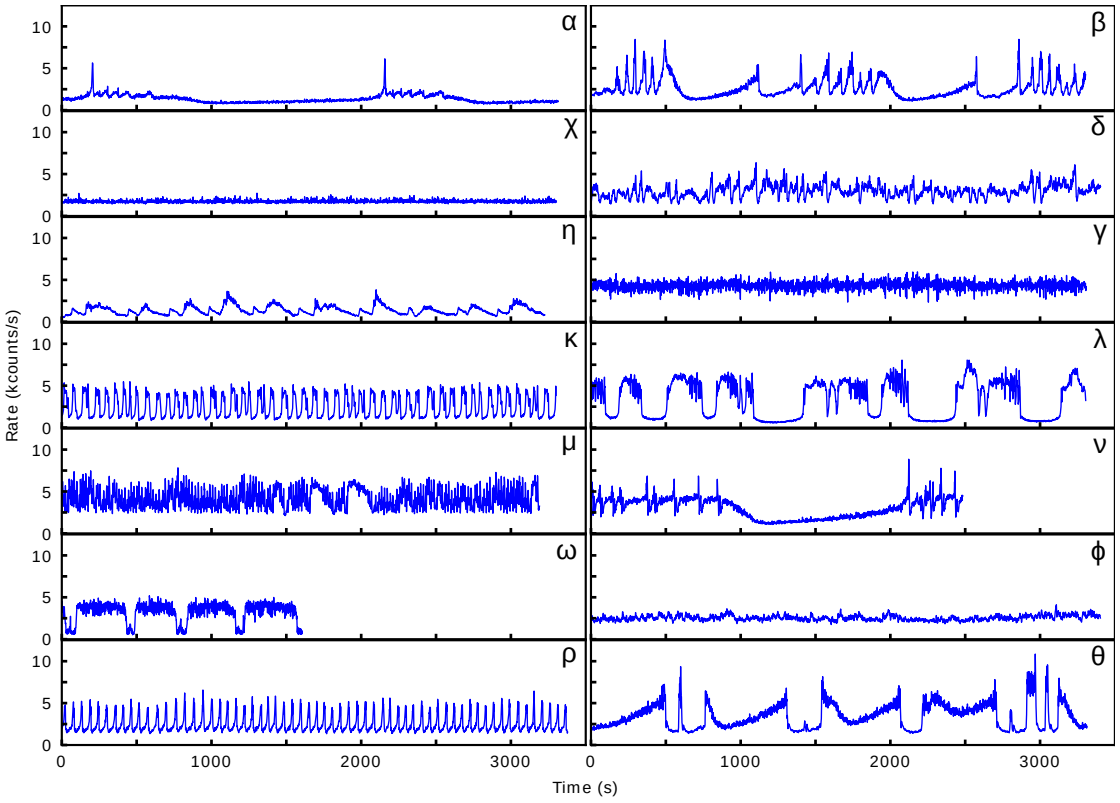


FIGURE 2.1: Examples of GRS 1915+105 X-ray light curves. Shown light curves have been classified according to the Belloni et al. system. Classifications are shown in the upper-right corner of each sub-figure. I use the curated set of classifications from Huppenkothen et al. (2017). Light curves have 1 second cadence.

resolution of 1 μs, and energy resolution of <18% at 6 keV. The high energy range (15-250 keV) was covered by the High Energy X-ray Timing Experiment (HEXTE, Gruber et al., 1996), comprised of two clusters of four detectors, with the total collecting area of 1600 cm^2. It had a field of view similar to PCA, the time resolution of 8 μs, and energy resolution of 15% at 60 keV. RXTE also carried an All-Sky Monitor (ASM, Levine et al., 1996), comprised of three scanning shadow cameras with the total collecting area of 90 cm^2, which covered 80% of the sky every 90 minutes, in the energy range of 2-10 keV.

2.1.2 GRS 1915+105

GRS 1915+105 is a galactic black hole X-ray binary system discovered in 1992 (Castro-Tirado et al., 1994), which shows an extraordinary complexity of X-ray flux variability. It was the only known source to exhibit such intricate patterns of behaviour, until the discovery of the black hole candidate IGR J17091−3624 (Kuulkers et al., 2003; Capitanio et al., 2006), which shares some of the same characteristics (Altamirano et al., 2011).

In an attempt to demonstrate that the complex behaviour of GRS 1915+105 is controlled by just a few simple physical variables, Belloni et al. (2000) constructed a classification system, which assigned observations of the source to one of 12 classes. Classification was based on the presence

of qualitative patterns in light curves and colour-colour diagrams of source observations. Work that followed had expanded the classification system to a total of 14 classes (Klein-Wolt et al., 2002; Hannikainen et al., 2003, 2005). This classification system is hereafter referred to as the "Belloni et al. system". Figure 2.1 shows an example of GRS 1915+105 light curve for each one of the 14 classes. Some classes show steady flux without any structured variability, other show periodic flares, dips or different types of periodic and aperiodic variability. There are both inter-class and intra-class variations in the amplitude of flux variability.

Highly complex patterns of light curve variability of GRS 1915+105 are commonly interpreted as being caused by a partial or complete disappearance of an observable innermost region of the accretion disc. Disappearance of the disc, in turn, is caused by thermal-viscous instability of the inner region of the disc (Belloni et al., 1997a,b). X-ray variability patterns corresponding to the 14 classes of source behaviour can repeat almost identically months and years apart. This suggests that the instability of the disc triggers a very specific and reproducible response (Belloni, 2001).

GRS 1915+105 was the first discovered stellar-mass source producing highly relativistic jets (Mirabel and Rodriguez, 1994). In this regard, it sparked great research interest, as it offered the possibility of studying coupled inflow-outflow processes of an accreting black hole, which unlike more massive active galactic nuclei, evolves in observable time scales (Fender and Belloni, 2004). Plasma ejections of GRS 1915+105 in the form of jets have also been associated with the instability of the accretion disc (Belloni et al., 2000; Nayakshin et al., 2000; Fender and Belloni, 2004), which supports the notion of disc-jet coupling.

Furthermore, Naik et al. (2002) found that intervals when the innermost region of the accretion disc is not observable (which are associated with variability class χ), are preferentially followed by classes showing periodic bursts, i.e. classes ρ and α. Therefore, following the changes of variability classes can improve our understanding of the evolution of the source over longer time scales, and it is an important method of probing the accretion dynamics, as pointed out by Huppenkothen et al. (2017).

Huppenkothen et al. (2017) conducted the first study of the whole set of GRS 1915+105 observations from RXTE/PCA using machine learning. They characterised the observations of the source and classified them according to the Belloni et al. system, using machine learning classification methods. They used features derived from power spectra, time series features extracted with an autoregressive model, and hardness ratio features derived from energy spectra.

2.1.3 Summary of this work

This section presents the classification of GRS 1915+105 observations using exclusively time series features derived from light curve data in an unsupervised manner, using a neural network. I choose not to use energy and power spectral features, found in the works of Belloni et al. (2000) and Huppenkothen et al. (2017), hence making my method more generalisable to other data sets,

sources, and energy bands. In principle, any type of univariate time series data can be analysed using this method.

In order to perform machine learning on light curve data, they need to be represented with vectors of constant length. It is a common approach to segment light curves to create time series of constant length prior to feature extraction. However, it is not obvious how to aggregate features of a light curve from a variable number of segments. I propose the "fingerprints" representation as a method of aggregating segment-wise time series features into a constant-length vector which describes a whole light curve of any length, in a manner that is interpretable by machine learning algorithms.

I train a neural network which performs dimensionality reduction of light curve segments, producing time series feature vectors of individual segments. I then perform clustering of features of individual segments and use the number of segments assigned to each cluster to construct the fingerprint representation, taking advantage of the fact that segments showing similar patterns of light curve variability tend to cluster together. I show that the fingerprint representation can be used to quantify the similarity of light curves and to perform machine learning classification.

This chapter contains the following parts: Section 2.2 describes the data preparation process, which starts with RXTE/PCA observations and produces a data set of light curve segments, a suitable input for a neural network. Section 2.3 provides details about the neural network and describes the process of dimensionality reduction of the data set, which results in the encoded data representation (time series features). Section 2.4 describes the process of cluster analysis of the encoded representation. Section 2.4.1 describes how Gaussian mixture models are used to identify the set of light curve patterns from the encoded data representation. Section 2.4.2 shows how the set of light curve patterns is used to construct the observation fingerprint. Section 2.4.3 shows how fingerprint can be used to assign light curves to classes of the Belloni et al. system, and Section 2.4.5 demonstrates how the fingerprint can be leveraged to refine the classification system in a data-driven way. Section 2.5 summarises the main results, discusses limitations of the presented approach, and lists some ideas for further research.

2.2 Data preparation

I retrieve all available RXTE/PCA[1] observations of GRS 1915+105 in Standard-1 format (0.125 second resolution light curves which combine all energy channels over the range of 2–60 keV). Extraction is limited to the most reliable counting array number 2 (PCU2). This yields 1776 light curves, which are subsequently re-binned. I generate two data sets: one where binning is performed at 1 second resolution, and another where binning is performed at 4 second resolution. Two data sets are generated because I fix the input size of the neural network to 128 data points

[1] https://heasarc.gsfc.nasa.gov/db-perl/W3Browse/w3browse.pl

TABLE 2.1: Parameters of the two data sets of light curve segment.

Parameter	4s data set	1s data set
Cadence (s)	4	1
Segment length (s)	512	128
Stride length (s)	8	10
# segments	468,202	474,471

due to GPU hardware constraints (see Section 2.3.2 for more detail), therefore the time resolution of the light curves is the main parameter which influences the amount and type of information that is provided to the network. On the one hand, short time resolution data can resolve fast variability structures of the X-ray source, but it cannot capture longer variability structures in a light curve segment of fixed size. On the other hand, longer time resolution data can capture more context within a light curve segment, but any fast variability structures are smoothed or lost. Hence, I choose to train two separate models on the two data sets and compare them in order to explore the effect of changing data resolution.

Similarly to Huppenkothen et al. (2017), I perform light curve segmentation in order to produce a data set of segments of standard length. Only the good time intervals from each observation are segmented, and the interruption periods of missing data are skipped. I choose the segment size of 128 data points, resulting in a segment length of 512 seconds for data with 4 second resolution. A moving window segmentation is performed with a stride of 2 data points (8 seconds), yielding a set of 468,202 segments derived from 1738 observations which satisfy the segmentation criteria. This data set of light curve segments is hereafter referred to as "4s data set". The same set of 1738 observations, binned to 1 second resolution, is segmented with a stride length of 10 data points, yielding 474,471 segments. This data set of light curve segments is hereafter referred to as "1s data set". Stride size is adjusted in order to make the number of segments as close as possible to the 4s data set. Table 2.1 lists parameters of the two data sets.

The time duration captured by the segments is not sufficient to contain the longest cycles of flux variability produced by the source. For example, some observations of the class α show intervals of quiescence which last ~1000 seconds, and are interlaced by periods of flaring which last ~500 seconds. However, the main goal of my study is not to classify individual light curve segments, but rather to construct a new, data-driven system of segment templates, which create classifiable observation signatures when grouped together with other segments from the same light curve. See Section 2.4.3 for an example of a classification experiment which illustrates the use of observation signatures (fingerprints).

A small stride size is selected in order to maximise the number of light curve segments available for neural network training. It also ensures that the model is exposed to the full range of phase shift of light curve patterns (Huppenkothen et al., 2017).

Light curve segments are independently standardised; count rate values of each segment are mean centred and scaled to units of standard deviation. Segments are standardised in order to decouple their shape and intensity information. Many of the patterns observed in the variability

of GRS 1915+105 repeat at various mean count rate levels. Therefore, standardisation of the segments is likely to cause segments showing similar shape patterns to align in the latent feature space extracted from the data. I also allow for the possibility that some shapes could be shared by several classes of variability, as described by Belloni et al. (2000), but at different intensities, and standardisation can make it easier to draw links between those cases.

The resulting data set of light curve segments, together with corresponding count rate uncertainty values (needed to calculate the reconstruction error, see Equation 2.1), is used to train a neural network. The process is described in Section 2.3.

Original count rate levels of the light curve segments are also important in the data analysis, so intensity information is added to the final feature set in the form of four descriptive statistics; mean, standard deviation, skewness and kurtosis, which are calculated from the distribution of count rate of each segment. These statistics make up one of the two sets of features used in cluster analysis in Section 2.4, and this set of four intensity features is hereafter referred to as "intensity features of light curve segments" (IFoS).

2.3 Feature extraction with a neural network

In order to address the problem of dimensionality reduction of light curve data, I propose a Variational Autoencoder (VAE) with LSTM cells within its encoder and decoder blocks. Autoencoders have been used in the past for the purpose of feature extraction from astronomical data (for example Naul et al., 2018). Here I use a variational variant of an autoencoder, because it introduces a Gaussian prior which applies a regularization effect on the produced latent space. Non-regularized autoencoders can map similar inputs to very different latent variables (Hinton, 2013), therefore regularization of the latent space is required for the down-stream clustering and merging of clusters discussed in Section 2.4.

2.3.1 Architecture of LSTM-VAE network

The network is built using Keras, an open-source neural network library (Chollet, 2015) with Tensorflow backend, and Python 3 programming language.

Figure 2.2 shows a visualisation of the proposed LSTM-VAE architecture. The purpose of each layer is as listed below.

- `Encoder_input` creates an instance of a tensor with dimensions of the model input, i.e. a sequence of 128 values.
- `Encoder_LSTM` is a layer of CuDNNLSTM cells. CuDNN stands for the CUDA Deep Neural Network library, which was developed by NVIDIA (Chetlur et al., 2014). The

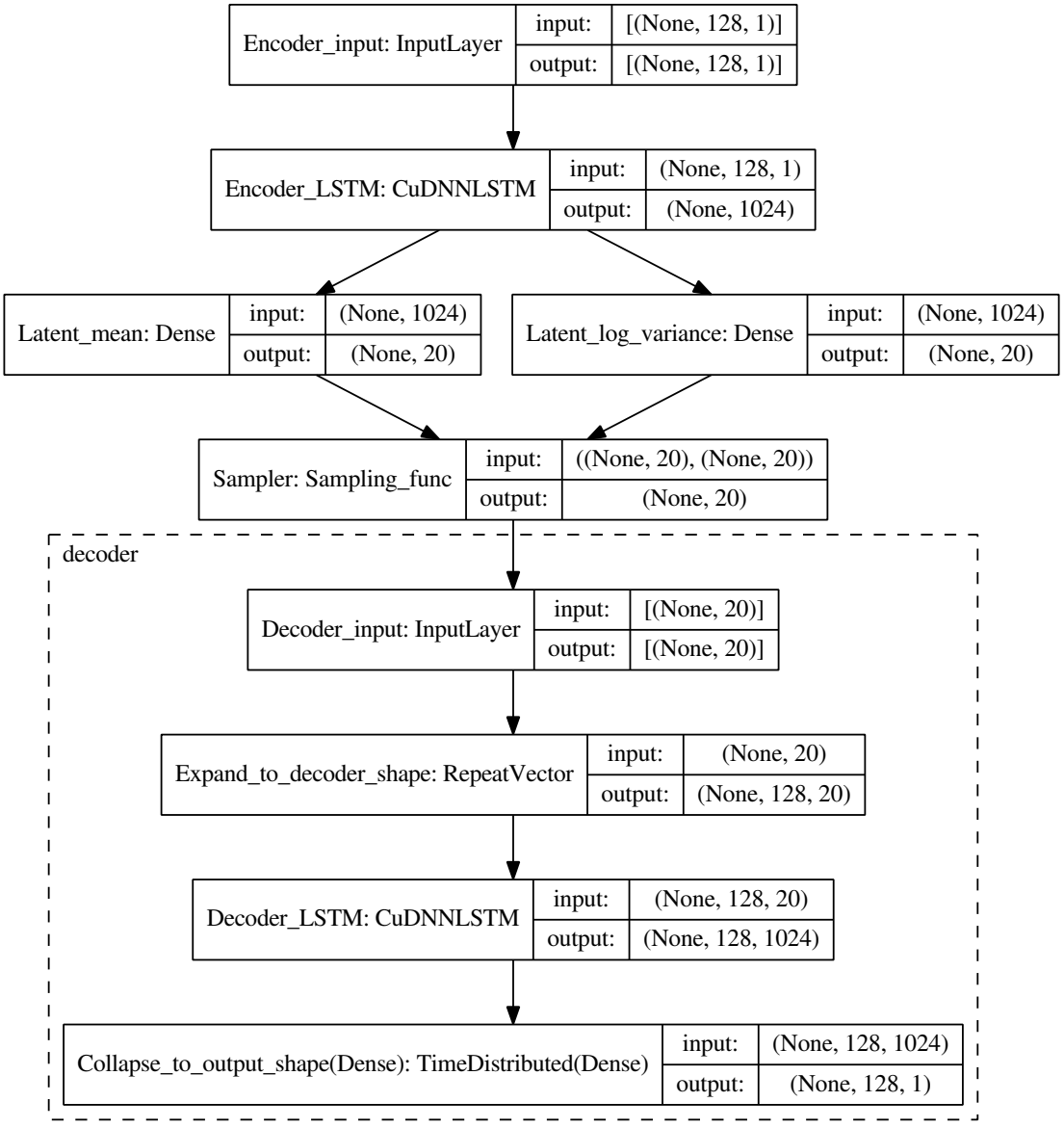


FIGURE 2.2: Architecture of the proposed LSTM-VAE model. The figure was generated using the KERAS utility `plot_model`. The left-most cell of each block contains the label assigned to the corresponding instance of a KERAS object, followed by class of KERAS model layer (except from `Sampling_func` which is a custom function; see Section 2.3.1). The right-most cells contain shapes of input and output tensors of the objects. Shapes are presented using the convention followed by KERAS.

library accelerates training of neural networks using a graphical processing unit (GPU). This layer consists of 1024 such LSTM cells, which are not interconnected, but perform recurrent computation on the input sequence, one data point at a time. Output from every point of the sequence is stored within the state of the cell and used as input of the next computation of the sequence. The output of this layer consists of the final state of the cells, produced after the entire sequence has been processed. LSTM cells are trained to extract informative variables from the data through the process of backpropagation of errors. Increasing the number of LSTM cells tends to improve the network's reconstruction loss (see below for the definition of reconstruction loss), and the number of 1024 is selected due to the GPU memory size constrain.

- `Latent_mean` and `Latent_log_variance` are two separate layers of fully interconnected neurons. Their purpose is to perform the dimensionality reduction of the 1024 variables extracted by the `Encoder_LSTM` layer. `Latent_mean` and `Latent_log_variance` each output 20 values (which is the dimensionality of the latent space). The first set of 20 values is used as the mean of the continuous latent variables, whereas the other set encodes their spread. In other words, `Encoder_LSTM`, `Latent_mean` and `Latent_log_variance` make up the Encoder block of the VAE, which maps the network input to the mean and (log) variance vectors. Increasing the number of latent variables tends to improve the network's reconstruction loss, but it also increases the dimensionality of the latent space and the computational complexity of downstream processes. I choose the number of latent variables based on the results of preliminary experiments, where increasing the number tended to improve the reconstruction quality, and the number of 20 reached the compromise between the reconstruction performance and the complexity of the latent space.

- `Sampler` generates random numbers from normal distributions whose parameters are set to the values of latent variable mean and variance. It is required in order to allow for deterministic treatment of the inherently probabilistic network during training (the so-called "re-parameterisation trick" (Kingma and Welling, 2014)).

- `Decoder_input` initialises the input tensor of the Decoder block of the model.

- `Expand_to_decoder_shape` replicates the values of latent variables to create sequences of the same length as the initial light curve sequences. In other words, each LSTM cell of the Decoder block receives values of the 20 latent variables at each iteration of the sequential computation.

- `Decoder_LSTM` layer is the Decoder counterpart of `Encoder_LSTM` layer of the Encoder. It also performs recurrent computation on the sequential input, but rather than processing a single sequence of variable values, it processes 20 sequences of constant values. This layer has 1024 cells, each producing a sequence of cell states from each iteration of recurrent computation.

TABLE 2.2: Summary of the LSTM-VAE training and fine-tuning. Validation loss stopped improving after the quoted number of training epochs. Training was stopped after 50 consecutive epochs without validation loss improvement. Best validation loss is shown.

Optimiser (rate)	# epochs	Loss	Data set
Adam (Default)	176	14.44	1s
SGD ($3 \cdot 10^{-4}$)	43	14.20	
SGD ($1.5 \cdot 10^{-4}$)	112	14.17	
SGD ($7.5 \cdot 10^{-5}$)	164	14.16	
Adam (Default)	209	107.63	4s
SGD ($3 \cdot 10^{-4}$)	5	103.67	
SGD ($1.5 \cdot 10^{-4}$)	4	102.55	
SGD ($7.5 \cdot 10^{-5}$)	12	102.02	

- `Collapse_to_output_shape` is a fully connected layer which combines the 1024 sequences from the LSTM layer into a single sequence of 128 data points. The output of this layer is the reconstruction of an input light curve segment.

2.3.2 Training and fine-tuning of LSTM-VAE networks

Both data sets are subdivided into training, validation and testing subsets. In order to ensure that the subsets are independent, segments derived from the same observation are only included in one of the subsets. In order to ensure completeness of the subsets, observations which have Belloni et al. system classifications available (Huppenkothen et al., 2017), are assigned to the subsets in a random, stratified fashion. At least one observation of each class is assigned to each subset, and the remaining observations are assigned to training, validation, and testing subsets according to the split ratio of 7/1/2. Since only two observations of η class are available at this stage, both observations are assigned to the training subset. Observations without Belloni et al. system classifications are randomly distributed between the three subsets, while accounting for the fact that observations contain variable number of count rate data points. The resulting training sets contain ~70% of total data points, validation sets contain ~10% of data points, and testing sets contain ~20% of data points.

Two LSTM-VAE neural networks with identical architecture are trained to compress light curve segments of 128 data points into 20 continuous latent variables, one network for each data set. Data from the training set is used to adjust the parameters of the networks, and the validation set is used to measure whether the adjustment improved the networks' ability to process previously unseen examples of data. The testing set is kept aside until all training and fine-tuning of models is finished.

I train networks using the KERAS implementation of Adam optimisation algorithm (Kingma and Ba, 2015), and fine-tune them using the stochastic gradient descend (SGD) algorithm. The `clipvalue` argument is set to 0.5 for both optimisers, which prevents numerical errors due to exploding gradients. Training is performed with the batch size of 1024 (number of light curve

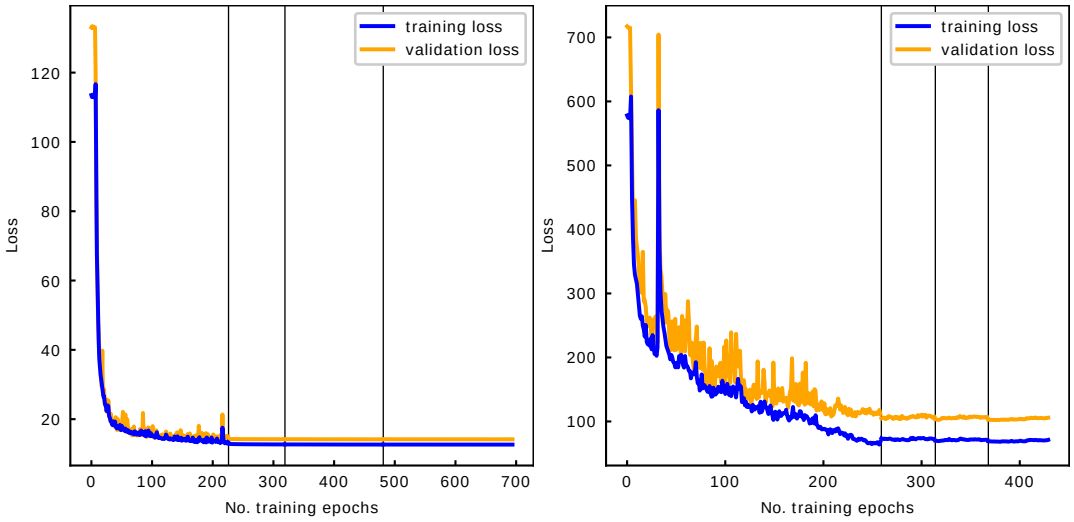


FIGURE 2.3: Loss curves resulting from the training of LSTM-VAE models on 1s data set (left) and on the 4s data set (right). Black vertical lines indicate that no improvement in validation loss in the last 50 consecutive epochs and change of the optimiser. See Table 2.2 for full training history.

segments propagated through the network simultaneously) for various numbers of epochs (i.e. complete passes through the training set). Training is terminated when the validation loss value does not improve for 50 consecutive epochs. See Table 2.2 for full training history, and Figure 2.3 for loss curves resulting from the training. Results suggest that satisfactory training results could be achieved using Adam optimiser only, because improvements caused by SGD with decaying learning rate are marginal.

I use an Nvidia Geforce Titan Xp GPU for network training. One epoch of training takes ~266 seconds of computation, resulting in the total of ~36.6 and ~17.0 hours of training for 1s and 4s data sets respectively, which could be reduced to ~13.0 and ~15.4 hours respectively, if fine-tuning with SGD optimiser was not performed.

The amount of GPU memory (12 GB for Titan Xp) sets a limit on the number of parameters of the network for a given size of data batch. The number of parameters depends on the size of propagated tensors, which in turn depends on the sizes of layers of the model (see shapes of the KERAS objects shown in Figure 2.2). In order to optimise the reconstruction performance within the constraints of the GPU memory limit, I maximise the number of LSTM cells, which increases the computational power of the network. Increasing the number of LSTM cells relative to the number of data points in the input layer (i.e. the segment length) tends to improve reconstruction performance. However, a decrease in the number of input data points means that less information about the context of variability patterns is given to the network, which shifts the focus of the analysis to short term patterns. The size of the training data batch could be reduced to allow for an increased number of parameters of the network, leading to improved performance at the expense of increased training time.

The output of the network consists of the reconstruction of an input light curve segment. Performance of the network is quantified using a loss function, which depends on the reconstruction error and a regularisation term. The loss function is a sum of a mean square error weighted by the data uncertainty (Equation 2.1) and Kullback-Leibler divergence (Equation 1.33). The weighted mean square error is defined as:

$$\mathrm{WMSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \hat{x_i}}{\sigma_i}\right)^2 \tag{2.1}$$

where N is the number of processed data points, x_i is the value of a data point i in the input light curve segment, $\hat{x_i}$ is the value of a data point i in the reconstructed sequence, and σ_i is the uncertainty of the value of i in the input light curve sequence. Uncertainty values are scaled with the same value of standard deviation as the count rate values of light curve segments.

The networks are trained to minimise the value of the loss function. Examples of light curve segments from 1s data set, together with their LSTM-VAE reconstructions, are shown in Figure 2.4. Examples for segments from 4s data set are shown in Figure 2.5.

2.3.3 Light curve feature extraction

The networks are used to perform dimensionality reduction on both data sets of standardised light curve segments (generated using the method described in Section 2.2). The network infers mean and variance of the 20 continuous latent variables from the data (i.e. outputs of `Latent_mean` and `Latent_log_variance` layers of the network correspond to the mean and variance parameters, see Section 2.3.1 for more details). The latent variables are a compressed representation of the network's input. For the purpose of further analysis, I do not sample from latent distributions, but take the means of latent distributions, which are representative of the position of each light curve segment within the latent space. The resulting set of latent variables is hereafter referred to as "shape features of light curve segments" (SFoS). Hence, the shape information of each light curve segment is represented by 20 SFoS values.

In order to assess how well SFoS describe the shape of light curve segments, I perform reconstruction of several segments. Figures 2.4 and 2.5 show reconstructions of light curve segments of each class of the Belloni et al. system. This set of segments demonstrates how the LSTM-VAE responds to a range of different light curve patterns found in the data sets. The model is able to reproduce the gross features of each segment, but it often does not account for fast count rate changes, which results in reconstructions which are significantly smoother than the input segments. This means that SFoS likely do not account for differences between segments lacking structured variability, where the major difference lies in the root-mean-square (RMS) deviation from the mean. Reconstructions of those segments would differ only in terms of the shape of random noise. For example, see segments of class ϕ and χ in Figure 2.5.

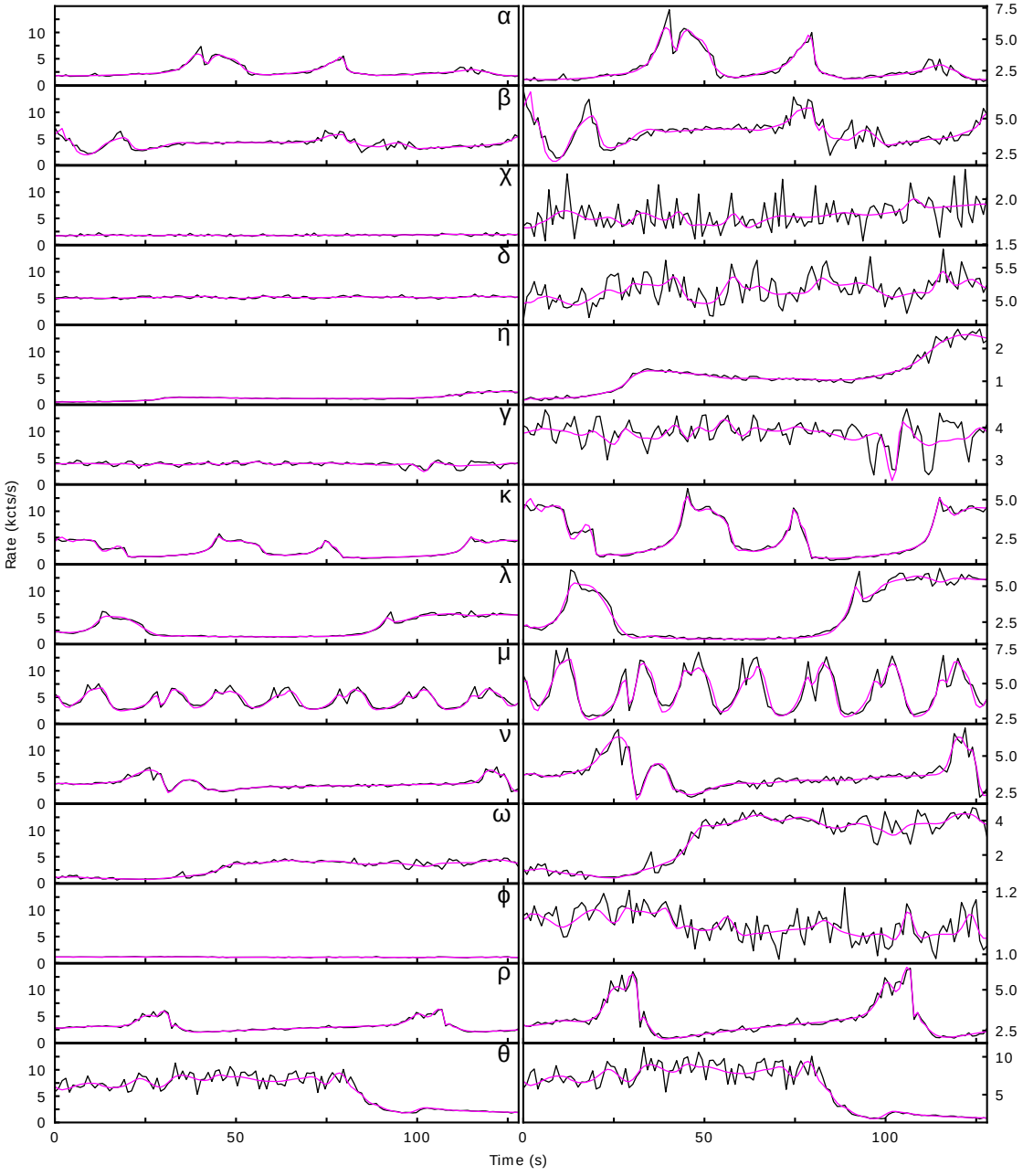


FIGURE 2.4: Examples of light curve segments from 1s data set. Segment reconstruction output of the LSTM-VAE is shown in magenta. Segments originate from observations that had been classified according to the Belloni et al. system. I use the curated set of classifications from Huppenkothen et al. (2017). Both columns show the same segments, but in the right column the range of the vertical axis is dynamic. All segments come from the testing subset of 1s data set, except for class η, which was included only in the training subset.

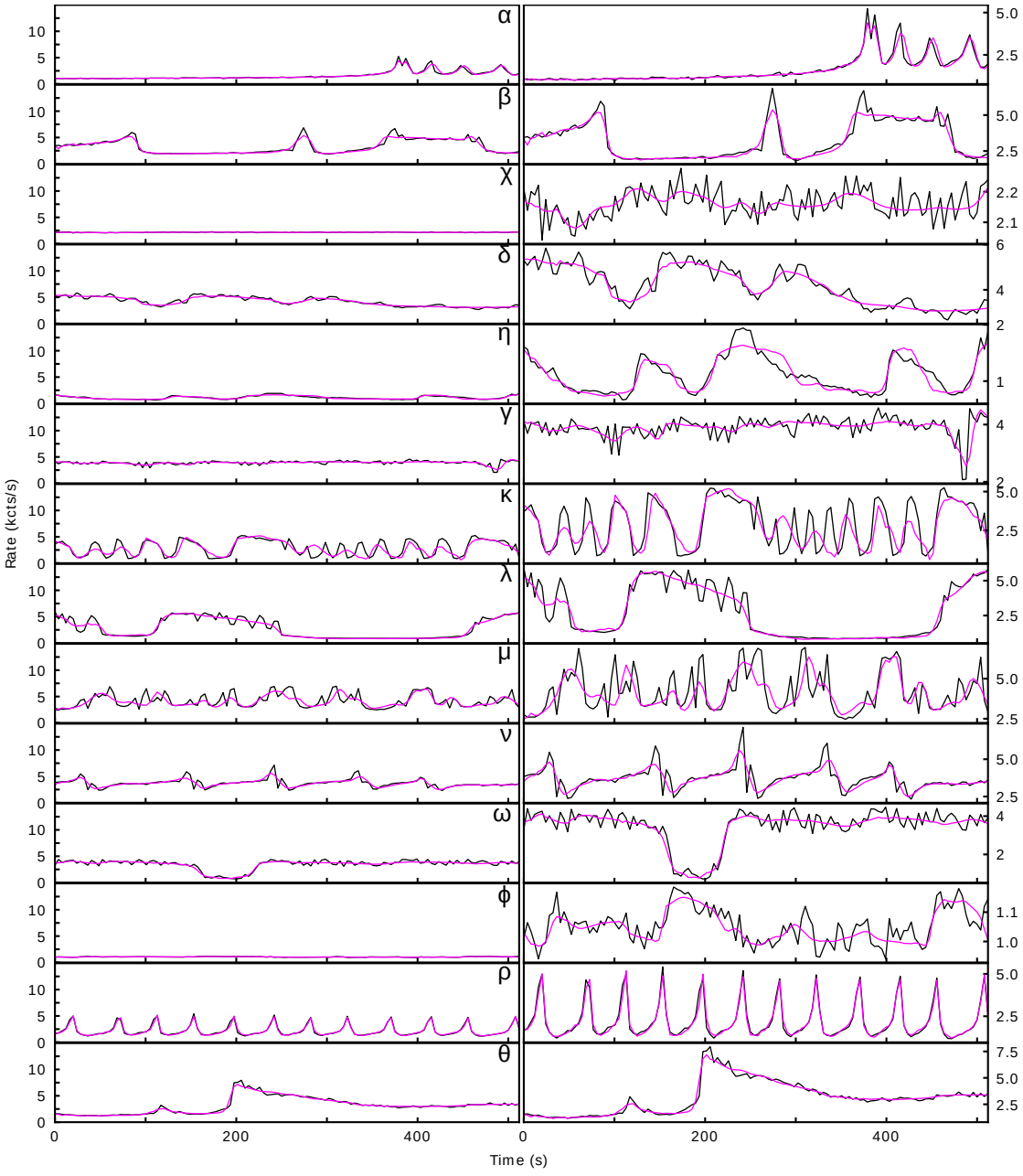


FIGURE 2.5: Examples of light curve segments and their LSTM-VAE reconstructions – the counterpart of Figure 2.4 for 4s data set.

In order to remedy this limitation, IFoS (containing information about count rate mean, standard deviation, skewness, and kurtosis of each segment) are used in the cluster analysis stage (Section 2.4) together with SFoS. Segments with indistinguishable shapes and dissimilar RMS can be distinguished based on their standard deviation values.

Figure 2.6 shows the distribution of IFoS values for segments of ϕ, χ and γ classes from the 1s data set. As expected, segments of class γ generally have larger mean and standard deviation values than the other two classes, which indicates that IFoS would allow for segments of class γ to be distinguished from classes ϕ and χ, even in segments where the characteristic "dip" of class γ is not observed.

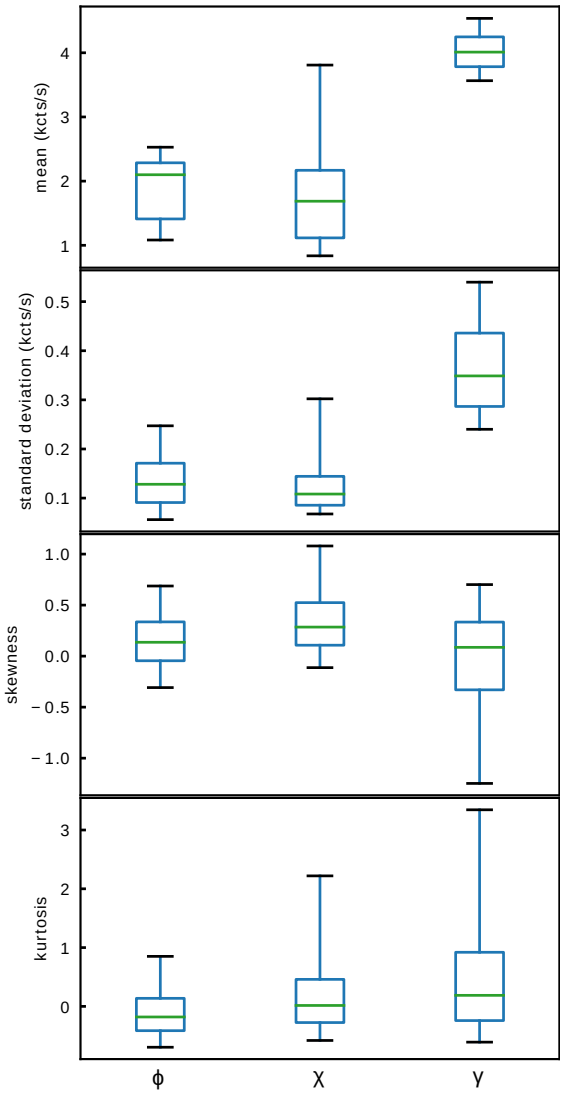
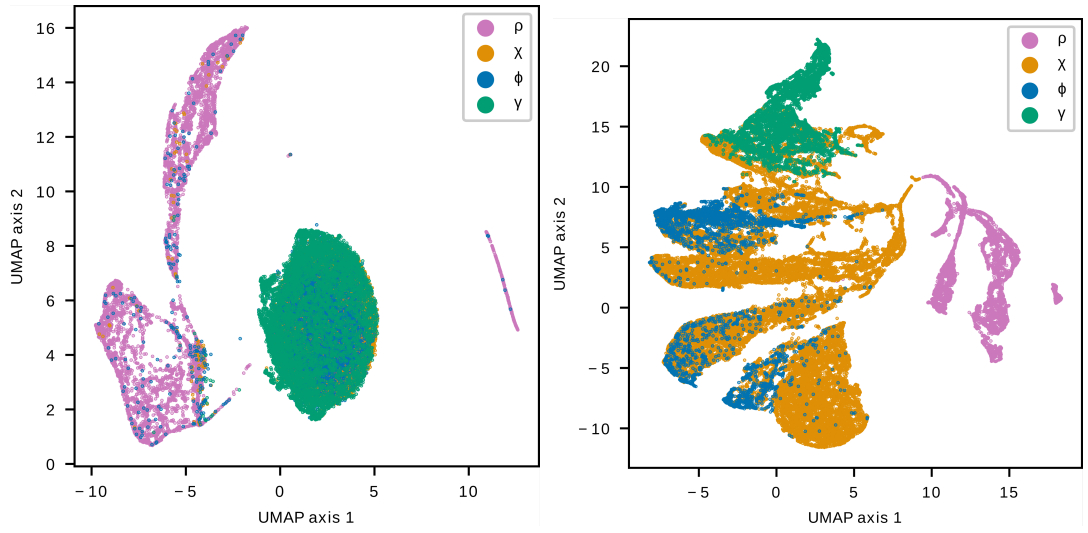


FIGURE 2.6: Distributions of the IFoS values from segments assigned to classes ϕ, χ and γ of the Belloni et al. system. Classifications come from the set curated by Huppenkothen et al. (2017). Four IFoS are the mean, standard deviation, skewness, and kurtosis of count rate values of each segment. The box extends from the first to third quartile, and the green line shows the median. Whiskers extend from the box to the 5th and 95th percentiles.

Furthermore, projections of SFoS and IFoS in Figure 2.7 show that IFoS could be used to distinguish classes ϕ, χ, and γ much more reliably than SFoS alone. The projection was produced using UMAP algorithm. In the projection of SFoS, data classified as ϕ, χ, and γ occupied the same region of UMAP space (in the central cluster of Figure 2.7a), indicating that those classes are mostly indistinguishable in the SFoS space. Data classified as ρ tends to occupy separate regions of the SFoS space. Class ρ is included in the projection to demonstrate that data associated with characteristic light curve shapes tends to be distinguishable in terms of its position in the SFoS space. IFoS projection uses the same subset of light curve segment data. Classes still show significant overlap in the IFoS projection, but intensity features can clearly provide meaningful information about the differences between classes ϕ, χ, and γ.



(a) SFoS (b) IFoS

FIGURE 2.7: UMAP projection of SFoS (left) and IFoS (right) of light curve segments from the 1s data set which have Belloni at al. system classifications of ϕ, χ, γ or ρ. Each point represents a light curve segment, and colour-coding shows their classification according to the Belloni et al. system.

2.4 Cluster analysis of generated features

2.4.1 Identifying the set of light curve patterns

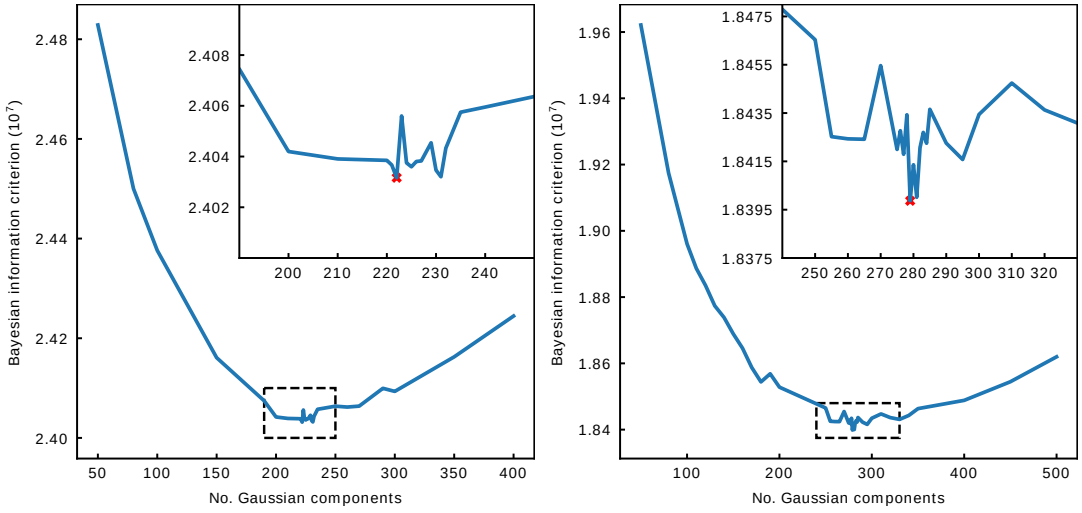
Sections 2.2 and 2.3 describe the process of feature engineering of four IFoS and 20 SFoS, which encode the shape and intensity information about light curve segments. This set of 24 features is hereafter collectively referred to as "shape and intensity features of light curve segments" (SIFoS). I generate two separate sets of SIFoS, one for the 1s data set, and one for the 4s data set.

In order to find the exhaustive set of light curve pattern templates which have been produced by GRS 1915+105, I perform clustering of the data in the 24 dimensional space of SIFoS. Clustering is performed with an implementation of Gaussian mixture model (GMM) included in the machine learning library SCIKIT-LEARN (Pedregosa et al., 2011). The algorithm approximates the probability distribution of the data using a set of multidimensional Gaussian components (see Section 1.4.4.1 for more details).

I explored the possibility of using density-based clustering algorithms DBSCAN (Ester et al., 1996) and OPTICS (Ankerst et al., 1999), and I find that it is difficult to fine-tune the hyperparameters of those algorithms in a way that would not lead to a large proportion of the data being rejected as noise. I find that GMM offers a much more straightforward method of assigning the data to clusters. Therefore, I focus on the discussion of the application of GMM in the remainder of this text.

I choose to use GMM to approximate the shape of the latent manifold in the SIFoS space, with the intention of merging of the Gaussian components which show significant overlap. This way

I use combinations of Gaussian components to account for the presence of any extended, curved, non-Gaussian structures in the latent space. I am assuming that those structures correspond to particular light curve patterns, and for each observation of the source, the relative amount of time the source spends showing those patterns allows us to determine the class of that observation.



(a) 1s data set (b) 4s data set

FIGURE 2.8: Bayesian information criterion of GMM as a function of the number of Gaussian components. I performed grid searches for the 1s data set (left) and the 4s data set (right). Figure insets are zoomed into the regions indicated by the black dotted line boxes, which contain the global minima. Minima were found at 222 Gaussian components for 1s data set and at 279 Gaussian components for 4s data set. Minima are marked by red crosses.

I perform a hyperparameter grid search to find the optimal number of Gaussian components for the GMMs in SIFoS space. Figure 2.8 shows values of the Bayesian information criterion (BIC) of those models as a function of the number of Gaussian components. BIC is a criterion commonly used to select the best model from a set of models fit to the same data set. The number of free parameters is one of the terms of BIC, so it penalises overly complex models. The model which produces the minimum BIC achieves the compromise between complexity and likelihood of the data (see Equation 2.2).

$$\mathrm{BIC} = k \cdot \log(N) - 2 \cdot \mathcal{L} \tag{2.2}$$

where k is the number of free parameters of the model, N is the number of samples of data, and $\mathcal{L}$ is the log-likelihood of the data.

The grid search for the optimal number of components of the GMM is the most computationally expensive procedure described in this chapter, barring the training of neural networks. The convergence time for this algorithm is highly dependent on the initial random state, but as an approximate point of reference, I inform that the fitting of the 222 component model to 1s data set and 279 component model to 4s data set require 1.96 and 2.91 CPU hours respectively, using

a CPU with clock speed of 3.40 GHz. Additionally, the fitting of 400 component models to 1s and 4s data sets require 3.42 and 3.80 CPU hours respectively.

I use the default set of GMM hyperparameters as implemented in the SCIKIT-LEARN library (Pedregosa et al., 2011), except for the number of components. I therefore fit full covariance matrices for each one of the Gaussian components. The computation time could be reduced at the expense of fit quality if a different type of covariance matrix was chosen. Convergence threshold and the number of iterations of the Expectation-Maximisation algorithm are other tunable parameters.

Grid searches indicate that the global BIC minimum for 1s data set is produced by a model with 222 Gaussian components, and for 4s data set with 279 components. I accept those numbers as the optimal numbers of components for GMM, however the stochastic nature of the algorithm could cause the numbers to change slightly if the grid-search was to be repeated. The set of clusters resulting from the assignment of light curve segment data points in SIFoS space to one of the Gaussian components of GMM is hereafter referred to as "Gaussian clusters".

Light curve segments showing similar type of variability patterns and count rate distribution are expected to produce similar values of SFoS and IFoS. Consequently, segments showing similar patterns are separated by smaller distances within the SIFoS feature space. Therefore, it is expected that Gaussian clusters contain homogeneous subsets of light curve segments, which are more similar to each other than to segments found in other Gaussian clusters.

Qualitative inspection of populations of the 222 Gaussian clusters of the 1s data set revealed that it is the case indeed. 42 Gaussian clusters contain recurrent flares, characteristic of light curve classes ρ and ν (see Figure 2.4 for examples). Table 2.3 shows a breakdown of the population of the clusters in terms of the classes of Belloni et al. system, and it shows that the Gaussian clusters containing recurrent flares are indeed dominated by classes ρ and ν. 69 Gaussian clusters contain segments with low RMS and no obvious patterns of variability, and their population consist mostly of classes χ and ϕ. Further 15 Gaussian clusters contain segments showing similar type of behaviour, but at higher average count rates, and the population of those clusters consist mostly of classes χ, γ and θ. Virtually every one of the remaining Gaussian clusters has some characteristic pattern; e.g. irregular flares, dips, negative or positive gradient, flaring followed by quiescence and vice versa. Few Gaussian clusters contain segments whose common patterns of variability are not immediately apparent upon visual inspection of a small random sample of segments, which can indicate that the number of Gaussian components of the GMM is too small.

In order to find the minimal, exhaustive set of light curve variability patterns, segments showing the same characteristic patterns should arguably all belong to one cluster. Gaussian clusters produced with the 222 and 279 component GMMs contain apparent degeneracies. The presence of very similar Gaussian clusters is caused by the limitation of the GMM, which approximates the probability density of the data set using multivariate Gaussian components. A single component cannot spread over a curved data manifold; only multiple components can approximate the

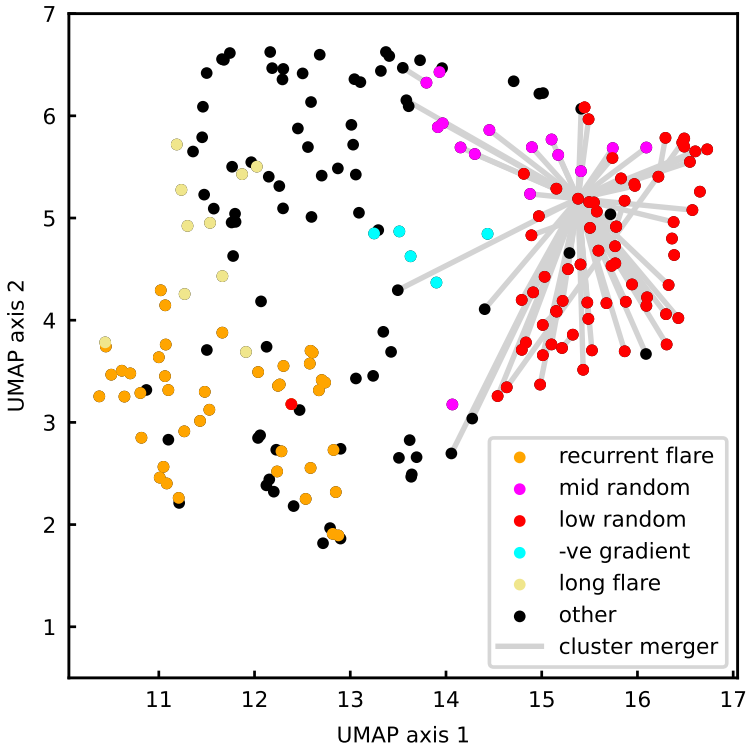


FIGURE 2.9: UMAP projection of the 222 means of Gaussian components of the GMM fit to 1s data set. UMAP reduced dimensionality from 24 to 2. Some points are colour coded, and common colours indicate Gaussian clusters containing light curve segments which show common characteristic variability patterns. As mentioned in the text, "recurrent flare" stands for behaviour characteristic of light curve classes ρ and ν, "mid-random" and "low-random" stands for the lack of apparent structured variability (only random noise) at low and medium-high mean count rates, "-ve gradient" stands for slow decrease of mean count rate, whereas "long flare" stands for irregular, long periods of flaring characteristic of classes κ and λ. All the remaining Gaussian clusters are labelled as "other", as their segments showed no obvious, common variability patterns. The population of each type of cluster in terms of the classes of Belloni et al. system is shown in Table 2.3. Grey lines indicate cluster mergers produced with the Mahalanobis distance threshold of 3.34. For every group of merged clusters, the lines are plotted between an arbitrarily chosen cluster and all the remaining clusters in the group.

curvature, as a series of locally flat sections. Light curve segments which show a similar type of variability pattern can vary in more than one SIFoS due to their non-linear interaction within the neural network model. Therefore, segments which, for example, vary in the frequency and amplitude of a similar type of pattern, can follow a curved manifold, and hence end up in separate Gaussian clusters. I address that in Section 2.4.3.

Figure 2.9 shows a two-dimensional UMAP projection of mean positions of GMM Gaussian components which are fit to the 1s data set. Some points are colour coded to indicate the characteristic patterns of light curve segments found in the corresponding Gaussian clusters. Segments in clusters labelled as "other" contain segments whose common patterns of variability are not immediately apparent upon visual inspection of a small random sample of segments. Colour coding is assigned through manual inspection of data found in Gaussian clusters, which is not a part of the proposed method of unsupervised data analysis. The purpose of this

TABLE 2.3: Percentage distribution of light curve segments across the classes of Belloni et al. system for each type of Gaussian clusters identified in Figure 2.9. Classes showing similar patterns tend to be clustered together, which is an indication of the meaningfulness of the SIFoS feature space.

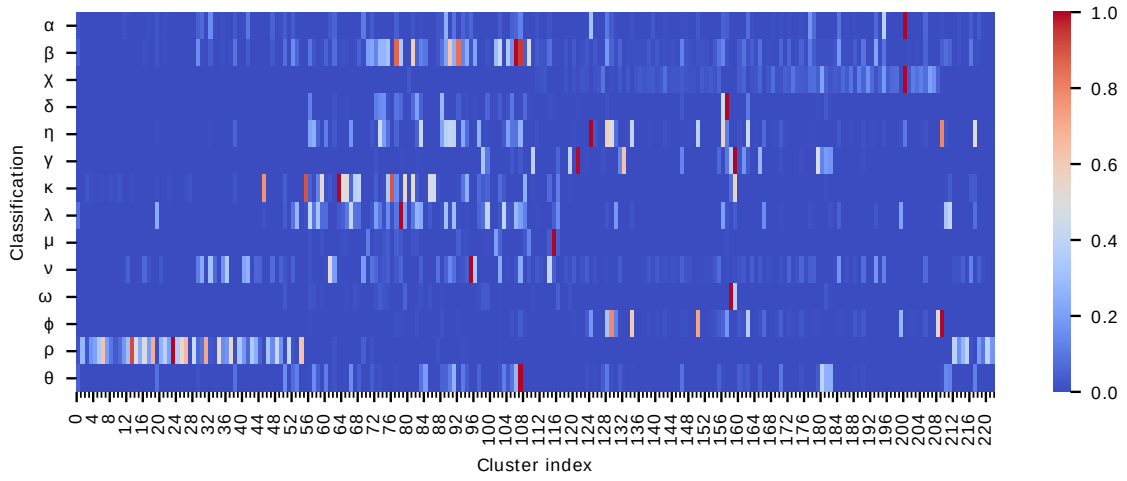
	recurrent flare	mid-random	low-random	-ve gradient	long flare	other
α	1.2	0.1	2.5	15.5	0.5	4.1
β	1	2.4	2.5	19.9	9.7	11.3
χ	0	46.9	80.4	0.1	0	22.6
δ	0	0.8	0	12.2	2.8	8.6
η	0	0	1.4	14.4	0	2.4
γ	0	27.5	0	0.1	0	8.2
κ	0.9	0.3	0	0	82.3	7.3
λ	0	0.1	0.4	3.4	1.1	4
μ	0.1	0	0	0.2	1.1	7.4
ν	5.9	0.9	1.3	2.8	0	3.5
ω	0	3.9	0	0.1	0	0.9
ϕ	0	1	11	4.8	0.1	1.9
ρ	90.9	0	0	0	2.3	3.4
θ	0	16.1	0.4	26.7	0.1	14.5

visualisation is to shows that Gaussian clusters sharing the characteristic patterns of behaviour tend to occupy similar regions of the SIFoS space.

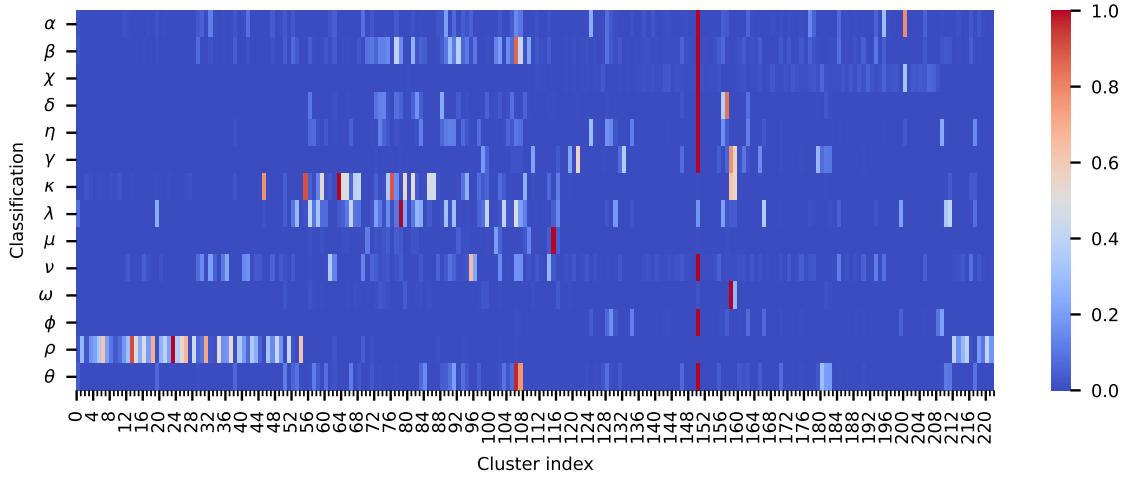
2.4.2 Relating the set of light curve patterns to the Belloni et al. system

In order to transform the two sets of clusters into a set of observation features, for each observation, I count the number of light curve segments assigned to each of the Gaussian components. This results in 222 values per observation in 1s data set and 279 values per observation in 4s data set. For each observation, I divide the counts by the sum of all counts for that observation. Feature vectors are independently normalised in order to reduce the impact of variance in the total number of segments extracted from each observation. This variance is caused by the fact that observations vary in total duration and the number of data gaps. Since feature vectors are normalised, they only contain information about the relative abundance of light curve patterns within the corresponding observation. I refer to such 222-vectors and 279-vectors as observation fingerprint, because they allow identification of distinct classes of light curve variability.

In order to showcase the usefulness of fingerprint representation of data, Figure 2.10a shows the subset of 1s data set which has been human-labelled according to the Belloni et al. system, in terms of the Gaussian clusters described in Section 2.4. Figure 2.10a shows combined fingerprint for each class of observation (fingerprint of observations of the same class were summed to produce the combined fingerprint). Rows of this heat map correspond to the 14 classes of observations, and columns correspond to the 222 Gaussian clusters of light curve segments. Particular cells of the heat map reflect the relative abundance of segments of a particular class which have been assigned to a particular Gaussian cluster. The heat map was



(a) 222 clusters



(b) 150 clusters left after mering

FIGURE 2.10: (a) A heat map showing the distribution of light curve segments across the set of 222 Gaussian clusters which were fit to the classified subset of 1s data set. (b) A heat map showing the distribution of light curve segments across the set of 150 clusters. The set of clusters results from the cluster merging procedure described in Section 2.4.3, applied to the set of 222 Gaussian components (illustrated in Figure 2.10a). The heat map was normalised row-wise. Colour indicates the relative abundance of light curve segments in the corresponding cluster. Clusters are ordered based on their proximity in the SIFoS space; this was determined using a hierarchical (single linkage) clustering algorithm. Observations' class labels come from Huppenkothen et al. (2017).

normalised row-wise, which means that high values indicate the Gaussian clusters which are most closely associated with a particular light curve class. This in turn indicates which light curve patterns are most abundant in light curves of a particular class.

The distinct appearance of the rows of the heat map indicates that the representation of light curves in terms of Gaussian clusters allows us to distinguish observations of different classes of the Belloni et al. system. This indicates that the light curve representation which employs the set of patterns could serve as a viable feature space for supervised classification algorithms.

TABLE 2.4: List of hyperparameters included in the grid-search of classification experiment described in Section 2.4.3. Hyperparameters criterion and max_depth belong to the random forest classifier. The possible values of criterion are "gini" and "entropy", which stand for Gini impurity and information gain respectively (see Section 1.4.4.3 for more details), and max_depth controls the depth of decision trees of the random forest. "Merge distance" refers to the Mahalanobis distance threshold used in the process of merging Gaussian clusters. I test 100 values evenly spaced between 1.5 and 5. Hyperparameter producing the largest F1 values for the two data sets are also listed.

Hyperparameter	Possible values	1s data set	4s data set
criterion	"gini", "entropy"	"gini"	"gini"
max_depth	None, 5, 10, 15, 25	5	15
merge distance	between 1.5 and 5	3.34	2.84

2.4.3 Classification of light curves using fingerprint representation

In order to test the usefulness of the Gaussian cluster representation in light curve classification, I train random forest classifiers (Breiman, 2001) to assign observations to the Belloni et al. system based on their fingerprint (note: I choose to disregard the subdivision of the χ class, because the main distinguishing feature of those subclasses is the position in the colour-colour diagram). I perform classification with the random forest classifier implementation included in the machine learning library SCIKIT-LEARN (Pedregosa et al., 2011) (see Section 1.4.4.3 for more details about the algorithm). I train separate classifiers for 1s data set and 4s data set.

In order to address the issue of degeneracy of light curve patterns described in Section 2.4.1 I merge Gaussian clusters separated by small distances. I use the Mahalanobis distance metric between the mean positions of Gaussian components. Mahalanobis distance is a distance metric used to measure the distance between a point and a distribution, while scaling the distance using the variance of that distribution. Mahalanobis distance is defined as:

$$D_{\mathrm{M}} = \sqrt{(\boldsymbol{u} - \boldsymbol{v})\mathbf{V}^{-1}(\boldsymbol{u} - \boldsymbol{v})^{T}} \tag{2.3}$$

where u and v are vectors whose separation is calculated, and V is the co-variance matrix.

The distance threshold between means of Gaussian components which are merged is one of the hyperparameters included in the grid-search during the training of random forest classifiers. Gaussian components are allowed to merge if the distance calculated for both co-variance matrices is smaller than the threshold hyperparameter. I discuss the effect of merging the Gaussian components in Section 2.4.4.

I use the training and validation data subsets (described in Section 2.3.2) for the training of random forest classifiers. I find that the label of observation 10258-01-10-00 from Huppenkothen et al. (2017) disagrees with preceding literature (Klein-Wolt et al., 2002; Belloni and Altamirano, 2013), therefore I change the label from λ to μ prior to classifier training.

Due to the limited number of labelled observations, I do not perform n-fold cross-validation, but instead I train the random forest classifier on 137 observations randomly sampled in a stratified

manner from the data set of 159 observations which were used for training and validation of the neural network, and use the remaining 22 for validation. I repeat this process 100 times for each combination of hyperparameters and find the mean of performance scores resulting from the 100 validation trials. In order to account for the class imbalance, the classifiers automatically adjust weights of each training sample to be inversely proportional to class frequencies. Table 2.4 lists hyperparameters included in the grid-search. The Scikit-learn (Pedregosa et al., 2011) implementation of the random forest classifier controls the number of decision trees in the random forest ensemble using the n_estimators hyperparameter. Therefore, I keep this number constant for all the grid-search classifiers, and set the hyperparameter to 1000 for the final classifier used on the testing data subset.

Hyperparameters min_samples_split and min_samples_leaf control the construction process of the decision trees, and I keep the values constant at 2 and 1 respectively. These hyperparameters control the splitting of the trees, and increasing the values would not allow for the classes with just a single observation to be separated from other classes. I set the class_weight hyperparameter to "balanced", which sets the weights of the classes to be inversely proportional to the class frequency in order to account for the imbalance in the number of observations of each class. Any hyperparameters of the random forest classifier which are not mentioned in the text were set to default values.

I use F1 and accuracy scores to measure the performance of classifiers. Both scores can take values in the range between 0 and 1, the higher, the better. Accuracy is the proportion of correct classifications out of all predicted classifications. F1 score is the harmonic mean of recall and precision of classifications of a single class:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{2.4}$$

where precision is the proportion of true positives out of the sum of all positives, and recall is the proportion of true positives out of the sum of true positives and false negatives.

Reported F1 scores are weighted averages across the 13 classes. The scores are weighted by the number of observations of the corresponding class in order to account for the class imbalance. In general, weighted F1 is a more reliable performance indicator, because accuracy can be easily biased when observations of one class significantly outnumber observations of other classes.

I find that the highest average validation performance scores for the 1s data set are weighted F1 of 0.814 ± 0.065 and accuracy of 0.854 ± 0.054, while the highest average validation scores for the 4s data set are 0.760 ± 0.068 and 0.810 ± 0.056 (reported uncertainty values are equal to one standard deviation calculated from the performance scores of the 100 validation trials). Therefore, I conclude that features derived from 1s data set perform better in the task of light curve classification. Hyperparameters producing the highest average validation scores are listed in Table 2.4 for both data sets.

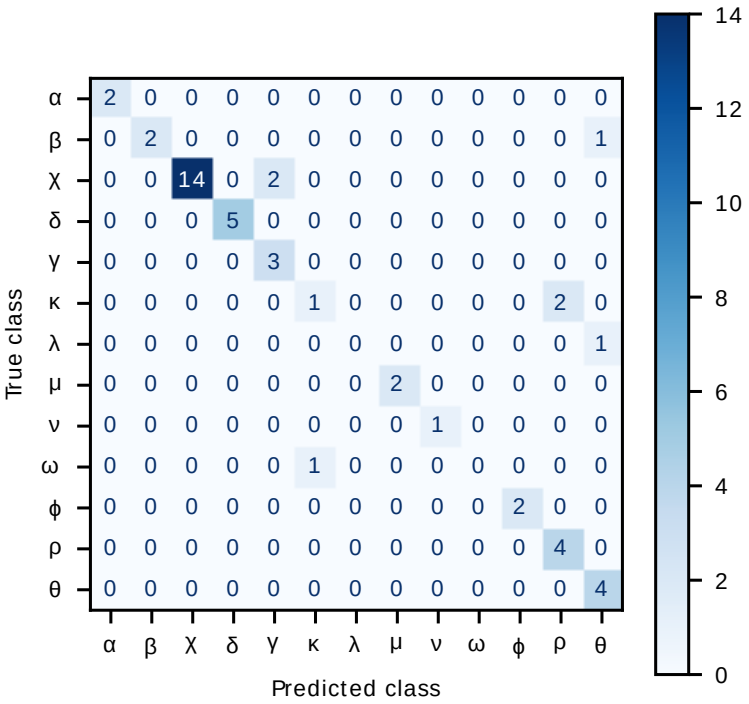


FIGURE 2.11: Confusion matrix showing classification results produced by the random forest classifier with optimal set of hyperparameters for 47 testing set observations. The matrix shows results for the lowest weighted F1 score out of the one thousand random initiations of the classifier. Weighted F1 and accuracy of this initiation are 0.834 and 0.851 respectively. The mean weighted F1 and accuracy performance scores across the one thousand initiations are 0.878 ± 0.027 and 0.894 ± 0.027 respectively

A random forest classifier with the optimal set of hyperparameters is trained on the observations of the training subset and tested on the testing subset of 1s data set, containing 47 observations. Random initiation of the algorithm is a significant cause of variance in the model performance, therefore training and testing is repeated one thousand times. The mean of weighted F1 and accuracy performance scores of those classifications are 0.878 ± 0.027 and, 0.894 ± 0.027 respectively. It should be noted that reported uncertainty values account only for variance caused by changes to the initial random state of the classifier.

Figure 2.11 shows the classification results with the lowest performance scores out of the set of 1,000 testing trials. I present the results of the test which achieved the lowest scores, because they reveal the most information about the low-confidence classifications. It appears that my classifications disagree the most with the human-assigned labels for classes that have few observations available. Therefore, it is likely that classification performance could improve given a larger amount of training data. Since the number of labelled observations is very small, the variance of test results is likely to be high. In order to test this, for each combination of hyperparameters shown in Table 2.4, I perform 100 tests where 69 test observations are randomly sampled in a stratified manner from the entire data set of 206 labelled observations (combined training, validation and test data subsets), and the remaining observations are used for training. Figure 2.17 shows the aggregated classification results of the 100 tests of the classifier with optimal hyperparameters. Those results seem to agree with the results shown in

Figure 2.11. The mean of weighted F1 and accuracy performance scores of those classifications are 0.877 ± 0.031 and, 0.896 ± 0.026 respectively. The results also show that classes which have only one training observation available (λ, ν and ω) tend to disagree with the human-assigned labels the most frequently.

From the test results presented in Figure 2.11, I examine each observation whose classification disagrees with the human-assigned labels. Observations of class β get assigned to other classes, most likely because of the complex behaviour of its light curves. Some patterns of class β can be seen in light curves of other classes (Belloni et al., 2000). Observation with ID 40703-01-35-01 belonging to class β (Klein-Wolt et al., 2002) is assigned to class θ. This observation contains many periods of missing data, and the good time intervals contain dips similar to the class θ. The observation also contains W-shaped intervals which resemble those characteristic of class θ. Furthermore, the classifier predicts that β is the second most probable classification of this observation (top three predictions are θ (37.1%), β (30.4%) and δ (6.1%)).

Two observations of class χ, 10408-01-42-00 and 40703-01-20-03, are assigned to class γ. Both observations show significantly higher count rate and RMS than an average χ observation, which is a possible cause of their classification. Class χ is the second most probable class prediction for both observations; the top three predictions for 10408-01-42-00 are γ (21.8%), χ (21.8%) and ϕ (14.8%), whilst for 40703-01-20-03 they are γ (21.9%), χ (21.9%) and ϕ (14.7%).

Two observations of class κ, 40703-01-24-00 and 40703-01-25-00, are assigned to class ρ. Both observations show fairly regular, sharp flares, similar to those characteristic to class ρ, however they are noticeably wider than the canonical ρ flares (see Figure 2.12 for the light curves of the two observations). Other than the similarity to ρ light curves, another factor influencing the classification of these observations is likely the heterogeneity of the light curve behaviour of observations labelled as κ by Klein-Wolt et al. (2002). More details about the ambiguity of κ classifications is provided in Section 2.4.5. Furthermore, class κ is the second most probable class prediction for both observations. Top three predictions for 40703-01-24-00 are ρ (16.0%), κ (15.5%) and χ (14.0%), whilst for 40703-01-25-00 they are ρ (15.7%), κ (15.3%) and χ (13.4%).

Observation 20402-01-36-01 belongs to class λ and is assigned to class θ by the random forest classifier. This observation shows behaviour which is very characteristic of class λ; it shows periods of flaring which alternate with low, quiet periods (see λ light curve in Figure 2.1 for an example). The likely cause of this class assignment is the fact that only one λ observation is included in the training data subset. The classifier predicts that λ is the second most probable classification; top three predictions are θ (30.1%), λ (20.8%) and κ (13.2%).

Finally, observation 40703-01-29-01 belonging to class ω is assigned to class κ by the classifier. This observation shows steady flux with no structured variability except a singe W-shaped dip, which is a very typical ω behaviour (see Figure 2.12 for the light curve). The classifier predicts that ω is the second most probable classification for this observation (top three predictions are κ (29.9%), ω (21.0%) and γ (16.8%)). One cause contributing to this classification is the fact

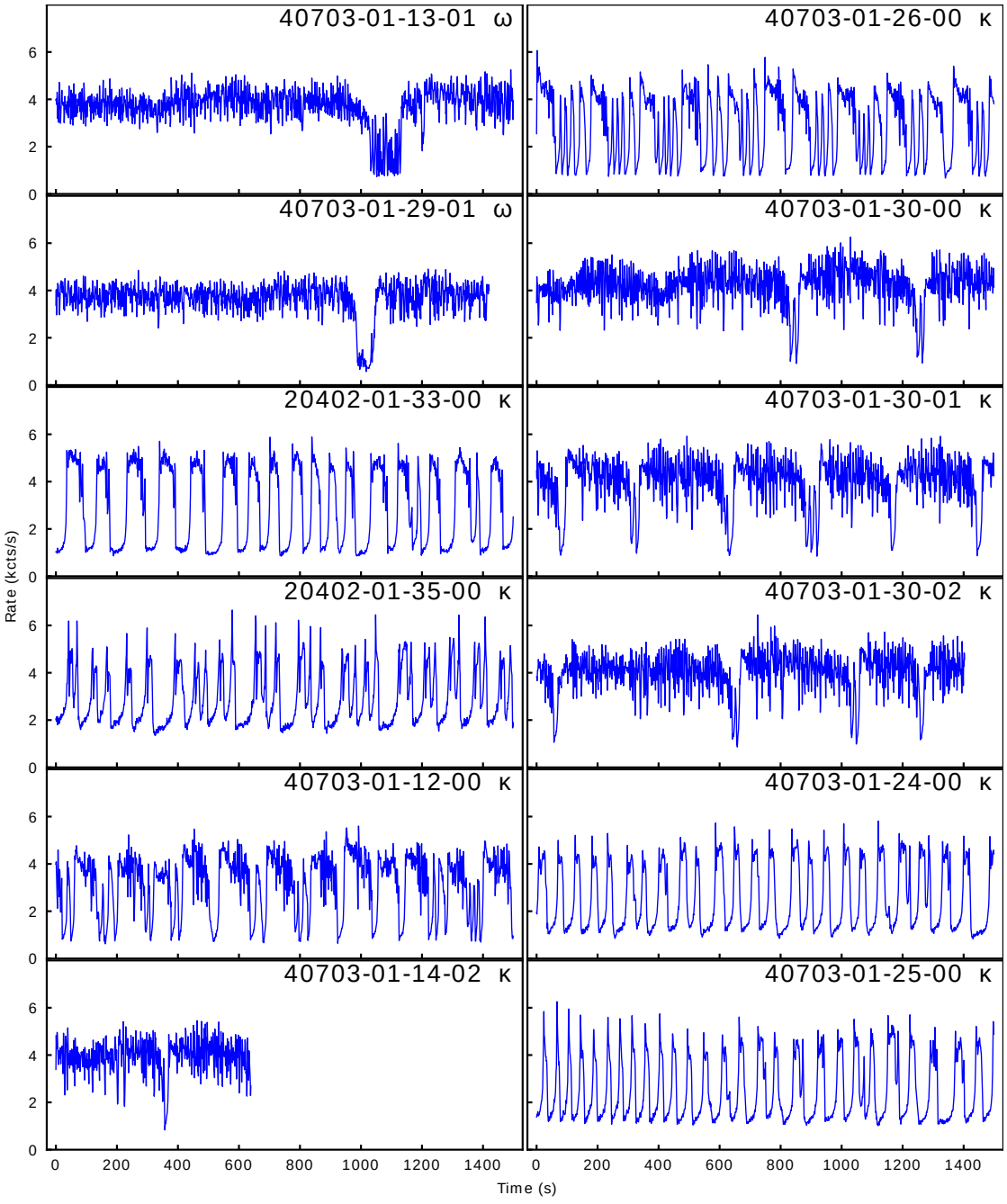


FIGURE 2.12: Light curves of ω and κ observations discussed in Section 2.4.5. Each sub-figure contains the first 1500 seconds of the light curve, or as much as is available in case of shorter observations. Each sub-figure contains the observation ID and classification from Huppenkothen et al. (2017). Figure 2.15 contains light curves which require individual time axes.

that only one labelled observation of class ω is available in the training data subset. However, another major cause becomes apparent upon inspection of ω and κ observations. Many of the observations labelled as κ by Klein-Wolt et al. (2002) show behaviour which very strongly resembles class ω. I discuss this issue in more detail in Section 2.4.5.

2.4.4 The effect of merging Gaussian clusters on the fingerprint representation

As discussed in Section 2.4.3 and shown in Table 2.4, the optimal classification result for the 1s data set is achieved with the Mahalanobis distance threshold of 3.34 for Gaussian cluster merging. This distance results in the reduction of the number of Gaussian clusters from 222 to 150. Grey lines in Figure 2.9 show which clusters are merged, where for every group of merged clusters, the lines are plotted between an arbitrarily chosen cluster and all the remaining clusters in the group. The merging results in nine composite clusters, eight of which consist of only two Gaussian components each, whilst one composite cluster consists of 65 components. All nine composite clusters contain only Gaussian components categorised "low-random" and "mid-random" (colour-coded as red and magenta points in Figure 2.9), and the "other" category with undefined variability (colour-coded as black). Out of the eight composite clusters made up of two Gaussian components, six of them contain only "low-random" components, one contained two "other" components, and one is made up of one "other" and one "mid-random" component. The largest composite cluster contained 46 "low-random" components, 11 "mid-random" components, and eight "other" components.

Gaussian components of the categories "low-random" and "mid-random" contain segments showing no structured variability. This type of variability can be found in the segments of multiple classes of observations; Figure 2.10b shows the distribution of light curve segments across the set of clusters after merging. The largest composite cluster, consisting of 65 Gaussian components (cluster index 150 in Figure 2.10b), contains the most abundant type of segment for most classes, including α, β, χ, δ, η, γ, ν, ϕ, and θ. The segments of those classes are distributed across multiple Gaussian components before merging, and they become more concentrated in a smaller number of clusters after merging, which is most the most pronounced for classes χ and ϕ. However, significant differences between the fingerprints of all classes are still noticeable. Since the distance threshold for cluster merging is adjusted to optimise the classification performance (as discussed in Section 2.4.3), it suggests that the large number of discrete clusters containing segments with random variability is not required to differentiate the classes.

2.4.5 Data-driven review of ω and κ classifications

Figure 2.13 shows the fingerprint representation of all ω and κ observations from Huppenkothen et al. (2017). There clearly exist at least two groups of κ observations. Six of the observations (40703-01-14-00/01/02, 40703-01-30-00/01/02) are much more similar to ω observations than the other κ observations. In order to assess the similarity of presented observations, I perform

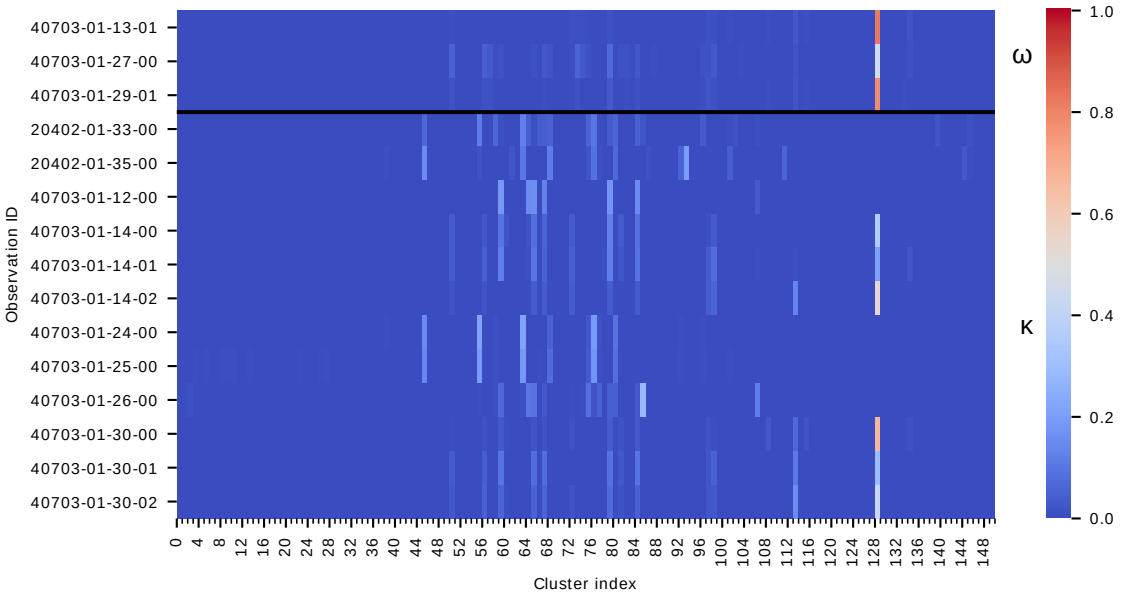


FIGURE 2.13: Fingerprint representation of ω (above black line) and κ (below black line) observations from Huppenkothen et al. (2017). Clusters are merged using the method described in Section 2.4.3, using the optimal Mahalanobis distance threshold of 3.34 (merged clusters are not plotted separately, unlike in Figure 2.10b). Colour indicates the relative abundance of light curve segments in the corresponding cluster.

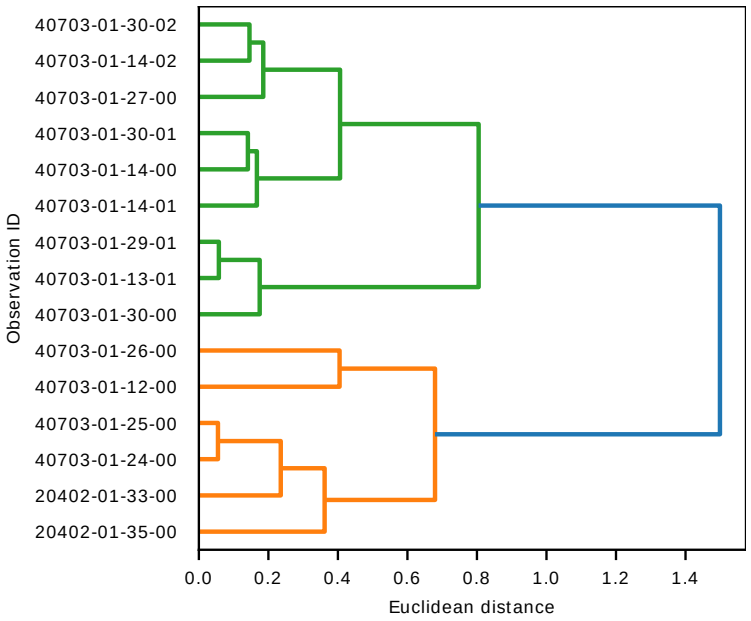


FIGURE 2.14: Dendrogram resulting from the hierarchical clustering of ω and κ observations based on their fingerprint, shown in Figure 2.13. Hierarchical clustering algorithm uses the ward method and Euclidean metric. Splitting of branches of the dendrogram at smaller values of Euclidean distance indicates that the observations in corresponding leaf nodes are more closely related.

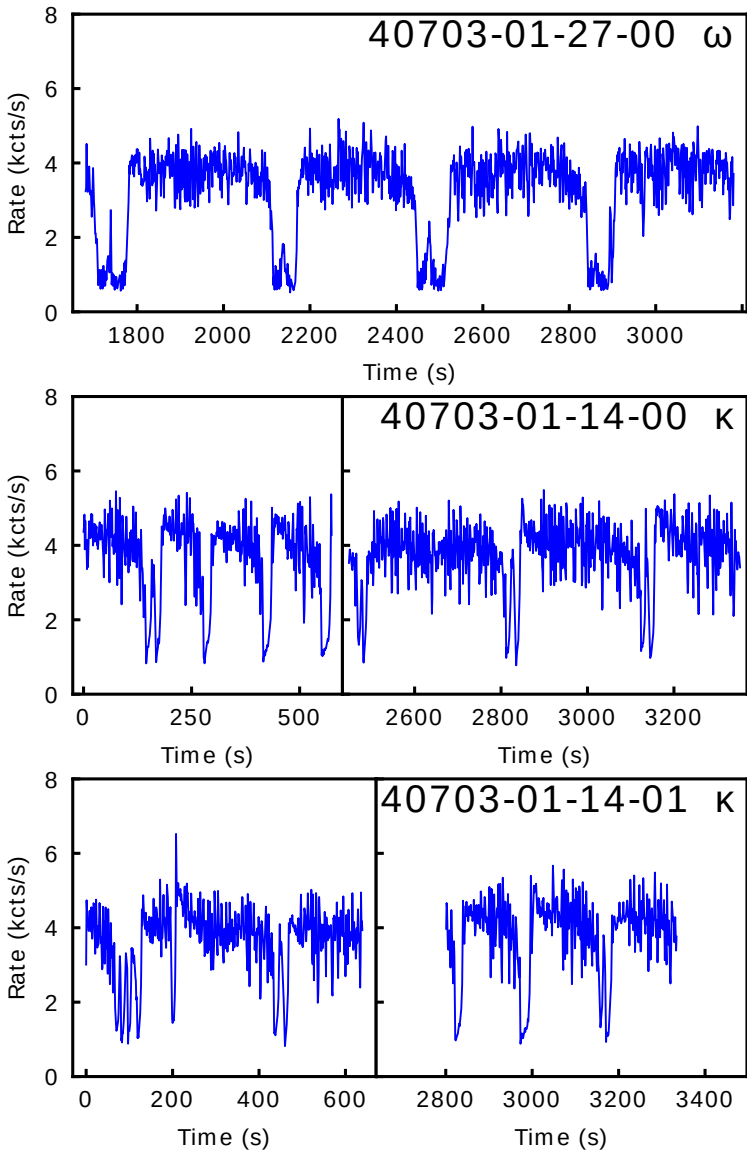


FIGURE 2.15: Light curves of ω and κ observations which are mentioned in Section 2.4.3. The top sub-figure contains only the second, longer good time interval of the observation. Middle and bottom sub-figures contain two good time intervals of their respective observations, and the data gaps are removed. Each sub-figure contains 1500 seconds of the light curve, or as much as is available in case of shorter observations. Each sub-figure shows the observation ID and classification from Huppenkothen et al. (2017). Figure 2.12 contains light curves of remaining ω and κ observations.

TABLE 2.5: Chronological list of RXTE/PCA observations of GRS 1915+105 during periods 51288-51306 MJD and 51394-51432 MJD. Classifications from Klein-Wolt et al. (2002) (Class. K), Pahari and Pal (2010) (Class. P) and Belloni and Altamirano (2013) (Class. B) are provided. IDs of observation whose classifications seem to be inconsistent are marked with †.

Observation ID	Date (MJD)	Class. K	Class. P	Class. B
40703-01-12-00 †	51288	κ	ω	ω
40115-01-01-00	51288	-	-	-
40403-01-07-00	51291	ω	ω	ω
40703-01-13-00	51299	γ	-	-
40703-01-13-01	51299	ω	ω	ω
40115-01-02-00	51299	-	-	-
40703-01-14-00 †	51306	κ	ω	-
40703-01-14-01	51306	κ	-	-
40703-01-14-02	51306	κ	-	-
40703-01-24-00	51394	κ	-	-
40703-01-25-00	51399	κ	-	-
40115-01-05-00	51406	-	-	-
40703-01-26-00 †	51407	κ	-	ω
40703-01-27-00	51413	ω	-	-
40703-01-27-01	51413	γ/ω	ω	-
40703-01-28-00	51418	ω	-	ω
40703-01-28-01	51418	-	-	-
40703-01-28-02	51418	ω	ω	ω
40115-01-06-00/01	51423	-	-	-
40703-01-29-00	51426	ω	ω	ω
40703-01-29-01	51426	ω	-	ω
40703-01-29-02	51426	γ/ω	-	ω
40703-01-30-00 †	51432	κ	ω	-
40703-01-30-01	51432	κ	-	-
40703-01-30-02	51432	κ	-	-
40703-01-30-03	51432	γ/κ	-	-

hierarchical clustering of observations in the fingerprint space. I use the hierarchical clustering algorithm included in the SCIPY package (Virtanen et al., 2020) and use the Euclidean metric and "ward" method, which uses the Ward variance minimisation algorithm and creates clusters of fairly regular sizes. (see Section 1.4.4.2 for more details about the algorithm).

Figure 2.14 shows a dendrogram resulting from the clustering of fingerprint of ω and κ observations (see Figure 2.12 and Figure 2.15 for their light curves). In the green branch of the dendrogram, observations classified as ω by Klein-Wolt et al. (2002) are clustered with the six κ observations mentioned above. These observations show semiregular dips with one or more re-flares. Observations 40703-01-13-01/29-01/30-00 clustered in the lower green sub-branch show lower frequency of dipping than the other observations, making them more alike to the canonical ω behaviour. Observations 40703-01-30-02/14-02/27-00 show a slightly higher dipping frequency, whilst observations 40703-01-30-01/14-00/14-01 show the highest frequency of dips in the green branch of the dendrogram. Similarity between fingerprint of κ observations 40703-01-14-00/01/02, 40703-01-30-00/01/02, and canonical ω observations is the result of the

similarity between the light curves of those observations. Based on this similarity, I suggest that those κ observations should be viewed as examples of an intermediate κ/ω state, with a major ω component.

Observations in the orange branch of the dendrogram show characteristic quasi-periodic and aperiodic flares of class κ. However, observations 40703-01-12-00/26-00 show a mix of flares with large and small width as opposed to 40703-01-24-00/25-00 and 20402-01-33-00/35-00, show a more consistent flare profile. Moreover, the presence of flares with greater width suggests an intermediate κ/ω state with a major κ component. Furthermore, Belloni and Altamirano (2013) indicate that observations 40703-01-12-00/26-00 belong to class ω, whilst both Pahari and Pal (2010) and Belloni and Altamirano (2013) indicate that observations 40703-01-14-00/30-00 belong to class ω, which shows that ambiguity of κ and ω classification exists in the literature.

Table 2.5 shows a chronological list of observations captured in the periods when the source was showing κ and ω behaviour, along with the classifications from Klein-Wolt et al. (2002), Pahari and Pal (2010) and Belloni and Altamirano (2013). Observations of class κ and ω were observed in close succession, which supports the notion of intermediate states between those classes. Based on the classification of observations in the two periods shown in Table 2.5, it seems that the source tends to transition between classes in the order $\kappa \rightarrow \omega \rightarrow \kappa$.

However, clustering results of the fingerprint representation should be interpreted with caution due to the geometry of the data sampling space (see Section 2.5 for more details).

2.4.6 Classification of 1024 second segments

I conduct an additional classification experiment with the goal of making it as comparable as possible to the work of Huppenkothen et al. (2017). The purpose of my work is primarily to introduce an unsupervised method of light curve data characterisation, whilst the primary objective of Huppenkothen et al. (2017) was to study GRS 1915+105, however both works use machine learning methods to perform classification on this source, so in spite of the differences, I believe that this experiment is informative.

I segment all the available RXTE/PCA observations of GRS 1915+105 in Standard-1 format to 1024 second segments and use the stride length of 256 seconds, yielding the total of 11028 segments, 2141 of which have human-assigned labels. I split this data between the training, validation, and testing data subsets in the ratio of approximately 50:25:25, ensuring that segments created from a single observation are not split between data subsets. Data is split in a stratified manner; at least one observation of each class is randomly assigned to each subset, and the remaining observations are distributed according to the ratio of 50:25:25. This results in 982, 592 and 566 segments in the training, validation, and testing subsets respectively. Further 6573 and 2314 segments of observations with no human-assigned labels are also assigned to the training and validation sets respectively.

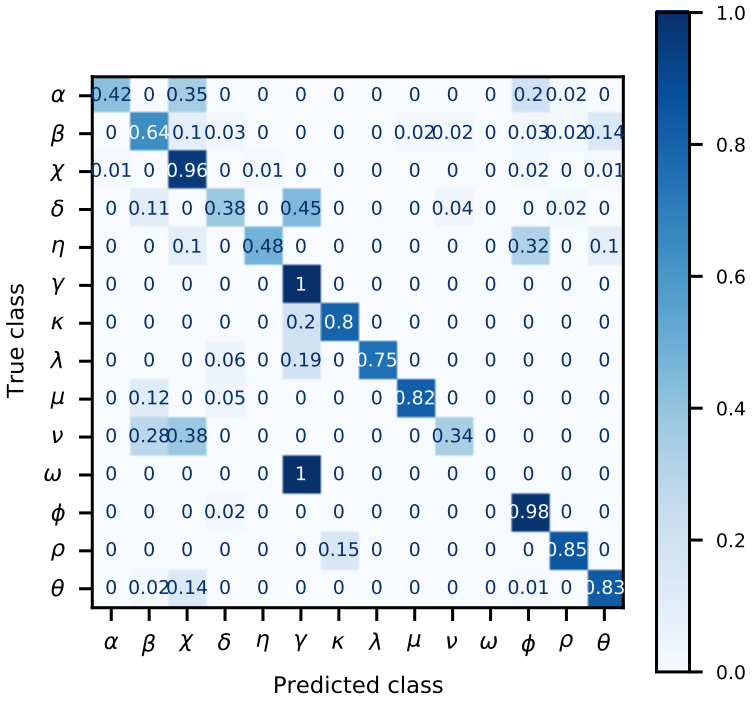


FIGURE 2.16: A row-wise normalised confusion matrix showing classification results for light curve segments of 1024 seconds. The numbers in each row are divided by the sum of numbers in that row.

In order to make the segments compatible with the LSTM-VAE network discussed in Section 2.3, I further subdivide the 1024-second segments into 64 non-overlapping segments of 16 seconds (which at the cadence of 0.125 seconds comprise 128 data points, as required by the network). I train the network using the training subset and evaluate the performance of the network after each epoch of training using the validation subset of data. I use the Adam optimiser with the `clipvalue` argument set to 0.5 to train the network. Training is stopped after 269 epochs, when the validation loss stopped improving for 50 consecutive epochs. I use the resulting network to encode the light curve segments and create 20 SFoS features.

I fit a Gaussian mixture model with 250 components to the SIFoS features of the training and validation subsets of data. I then perform a grid search of hyperparameters listed in Table 2.4, where for each combination of hyperparameters I train a random forest classifier on the training set and I evaluate the prediction on the validation set. I find that the highest weighted F1 and accuracy performance scores of 0.834 and 0.851 respectively are produced using the merge distance of 3.798, criterion of "entropy", and max_depth of 5.

I test the classifier with optimal hyperparameters on combined validation and testing data subsets, resulting in weighted F1 and accuracy performance scores of 0.811 and 0.822 respectively. The accuracy of this classification is ca. 10 percentage points lower than that of Huppenkothen et al. (2017). Their method is simpler and more accurate, but it requires a larger variety of data features, i.e. power-spectral and hardness ratio features. Therefore, depending on the available features, it might be the preferable method for the task of classification of 1024 second segments.

Including those features in my method could result in a performance improvement, which is a possible avenue of future investigation.

The possible avenues for improvement of both methods include the training on a larger amount of labelled data, and the reduction of ambiguity in the labelled data. Ambiguity is caused by the fact that the classes of the Belloni et al. system are not fully exhaustive and mutually exclusive. This is discussed further in Section 2.5.

Figure 2.16 shows the confusion matrix of the predicted labels against the human-assigned labels. Even though most of the segments are classified in agreement with the human-assigned labels, the drop in performance is significant when compared to the other experiments shown in this work (Section 2.4.3 and Section 2.4.7). In particular, all the ω segments are classified as γ, which is most likely caused by the noise in the data. The data with cadence of 0.125 second contains a significant amount of random noise, and the segmentation of data to the length of only 16 seconds followed by standardisation, results in segments which are seemingly featureless, with occasional peaks and troughs beyond the noise level. The characteristic dips of class ω were not preserved in this representation of data. In conclusion, re-binning of data to cadence longer than 0.125 seconds prior to LSTM-VAE training is likely to be beneficial for the performance of my method. Additionally, when light curves are segmented to 1024 seconds prior to classification, I cannot exploit the main advantage of the "fingerprint" representation, which aggregates features of long light curves. Light curves of classes whose cycles of variability evolve over longer time than 1024 seconds can yield segments which appear significantly different, depending on the phase of a variability cycle.

2.4.7 Classification test with all available data

Figure 2.17 shows a summary of the classification results for the 100 tests where 69 test observations are randomly sampled in a stratified manner from the entire data set of 206 labelled observations (combined training, validation and test data subsets), and the remaining observations are used for training (as described in Section 2.4.3). The mean of weighted F1 and accuracy performance scores of those tests are 0.877 ± 0.031 and 0.896 ± 0.026 respectively. Observations of the majority of the classes are classified in agreement with the human-assigned classifications >80% of the time, however, observations of classes λ, ν and ω are usually assigned to other classes. These classes are affected by very small sample size (each has only one training observation), which limits the statistical significance of their classification results.

2.5 Discussion and summary

In this chapter, I introduce a data-driven method of light curve feature extraction and test the utility of resulting features by conducting a set of supervised multi-class classification experiments, using a set of human-labelled observations. Light curve classification of data in 1 second

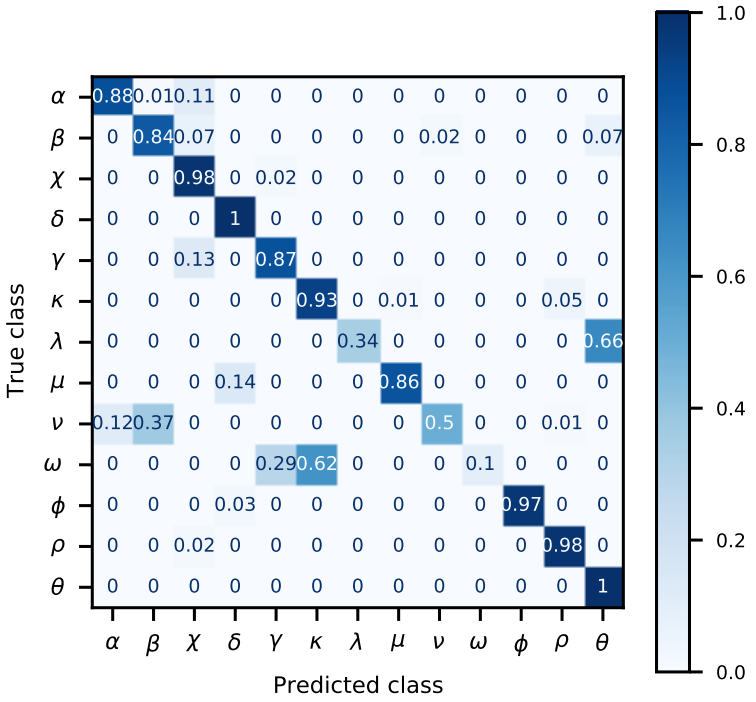


FIGURE 2.17: A row-wise normalised confusion matrix showing classification results of 100 test trials where 69 test observations are randomly sampled in a stratified manner from all the available 206 labelled observations, as described in Section 2.4.3. Rows of the matrix correspond to human-assigned classes (assumed true class), columns correspond to the classifications predicted by the classifier, and the values in the matrix represent the number of observations. The numbers in each row are divided by the sum of numbers in that row.

resolution resulting in a mean weighted F1 score of 0.878 suggests that the proposed method of unsupervised feature extraction is capable of producing features which represent light curve data in a meaningful way.

One possible application of the method is the unsupervised exploration of the data space. The shape of the latent data manifold encodes information about modes of activity of the source and its evolution between them. The analysis of correlation between SIFoS and various power-spectral and energy-spectral features could be performed in order to study the links between latent variables, the occurrence of different variability patterns and the physically interpretable parameters. For example, broad-band power spectral shape could help to link variability patterns with the geometry of the emitting regions (Heil et al., 2015), presence of high frequency quasi-periodic oscillations could link the patterns with the orbital motion of matter in the accretion disc (see Motta, 2016, for a review), and radio flux could link them with the emission of relativistic jets (Belloni et al., 2011).

Work presented in this chapter shows that dimensionality reduction of the data set, followed by clustering of observations in this reduced space could be a way to derive a set of classes of source behaviour, which avoids the biases of human characterisation and annotation of data. Furthermore, I find that Gaussian component merging based on the Mahalanobis distance between them

can help to reduce the problem of cluster degeneracy caused by the limitations of GMM, which requires multiple components to follow curved data manifolds.

The Belloni et al. system of classification discussed in this work is not comprehensive, and to some degree, it is arbitrary. As Belloni et al. (2000) point out, their classification system was not intended to exhaustively list mutually exclusive modes of behaviour. This creates a problem for any classification effort that is based on this system. A smaller number of classes could be chosen, because some classes show behaviour that arguably lies on the same continuum. One example of such a continuum could be followed by classes λ and κ, which show similar behaviour at slightly different time scales (Belloni et al., 2000). Furthermore, my work shows that classes κ and ω can show very similar light curve behaviour and possibly lie on a continuum as well. A larger number of classes could just as well be required, as it is possible that there exist additional patterns of behaviour that have not been characterised yet (similarly to classes ω and ξ (a.k.a. η) which were added to the Belloni et al. system later than the original 12). Transitions between classes are observed, so ambiguity in classification of observations cannot be avoided.

Taking these considerations into account, it is very difficult to ensure the accuracy of classification for large sets of unknown data, which cannot be standardised to adhere to the assumed classification system. Performance of any supervised classification algorithm greatly depends on the definition of the classes of observations, and any ambiguity is going to affect this performance. Therefore, clustering of the data set can help in deriving a data-driven set of observation classes, which helps to avoid the biases of human characterisation and annotation of data.

However, further work in this direction will need to address the problem of clustering of compositional data. The fingerprint representation uses vectors of fixed size, whose values sum up to a positive constant. This is a type of compositional data, and as such it is constrained to the geometry of a simplex (Aitchison and Egozcue, 2005), which means that results of clustering of raw compositional data are often unreliable. Clustering methods applied to compositional data should take into account the need for prior transformation of the data into an unbounded, Euclidean space. This issue is an open area of research, and it is further complicated by the presence of numerous zero values in the fingerprint compositional data (Aitchison et al., 2000; Martín-Fernández et al., 2003). Study of the appropriate transformation methods for compositional data goes beyond the scope of my work, therefore results of fingerprint clustering should be interpreted with caution.

Since the proposed feature extraction method is easily generalisable to different types of time series data, there exists a range of possible applications for the proposed feature extraction pipeline. A similar type of analysis is possible for sources other than GRS 1915+105, and in principle any energy band of light curves.

Fingerprint representation could be used to test whether the phenomenology observed in GRS 1915+105 is also present in other sources. For example, similarity in the variability of IGR J17091−3624 (Court et al., 2017), and the Rapid Burster, MXB 1730-335 (Bagnoli and In't Zand, 2015) could point to accretion physics which are independent of the nature of the accretor.

Pattnaik et al. (2021) attempted machine learning classification of compact objects in low mass X-ray binaries based on energy-spectral features, and found that fairly accurate classification was possible. The presence of a black hole or neutron star in the binary system can have a significant impact on the physical interpretation of the observed phenomenology. Classification of compact objects using the fingerprint representation of X-ray variability patterns is a possible subject of future work. However, care must be taken to select sources in the hard state, which show structured X-ray variability.

Derivation of fingerprints which represent the set of light curve patterns observed in a fixed amount of time could be the basis of a live monitoring system, which would alert the user about changes in the behaviour of the source. This could involve classifying observations using a known system of classes, but it could also involve the task of outlier detection, where the position of an observation would be tracked within the encoded feature space. Observations producing feature vectors which fall in sparse regions of the feature space would indicate an anomaly.

The main requirement of the proposed feature extraction method is that the light curves must be evenly sampled in time. This requirement is satisfied by data similar in nature to the pointed observations of RXTE/PCA. This includes data captured by the X-ray Telescope aboard the Neil Gehrels Swift Observatory (Swift) (Burrows et al., 2005) or X-ray Timing Instrument aboard the Neutron Star Interior Composition Explorer (NICER) (Gendreau et al., 2016) etc. High-speed optical telescopes ULTRACAM (Dhillon et al., 2007) and HiPERCAM (Dhillon et al., 2016) also produce light curves which are evenly sampled over the time of an exposure, and could be analysed using the method I propose. However, the amount of data produced by ULTRACAM and HiPERCAM is not large enough to justify the use of machine learning analysis. The Optical Timing Camera (OPTICAM) (Castro et al., 2019) will produce a larger amount of data of similar nature, and the proposed method of feature extraction could be appropriate for their analysis.

The proposed method could be used to characterise long-term light curves captured by all-sky X-ray surveys, like those performed with the Gas Slit Camera aboard MAXI (Matsuoka et al., 2009), the Burst Alert Telescope aboard Swift Gehrels et al. (2004) or the All Sky Monitor aboard RXTE, provided that light curves have regular time bins. Interpolation could also be performed if needed.

Pursiainen et al. (2020) interpolated ca. 30,000 light curves of the Dark Energy Survey Supernova Programme using Gaussian Processes, and increased the cadence from the average of 7 days to a constant 0.5 day cadence (see also Wiseman et al., 2020). Light curves generated by several ground-based surveys could be made compatible with my feature extraction method using such interpolation techniques. Those surveys include LSST, ZTF and Asteroid Terrestrial-impact Last Alert System (ATLAS) (Heinze et al., 2018). Optimisation of the type and size of kernel used in Gaussian Processes interpolation would need to be performed by the user prior to feature extraction, and the choice of parameters would depend on the nature of data and the scientific goal.

In addition to the requirement of having even sampling in time, light curves should not have many gaps which cannot be interpolated over, in order to be suitable for analysis using the proposed feature extraction method. During the light curve segmentation stage, segments are extracted using a moving window method, and segments which straddle over data gaps are discarded. Therefore, light curves must have few gaps to allow for a choice of segment size which is large enough to encompass time-scales which are relevant for the variability of the analysed data.

There are several limitations to this work. Figure 2.6 reveals that classes ϕ and χ are similar in terms of IFoS. The only significant differences are (1) that mean count rate values of segments of class χ span a larger range than the segments of class ϕ, and (2) that segments of class χ can have more positive outliers in terms of kurtosis. Figure 2.7 shows that the combination of IFoS contains enough information to distinguish many cases of those classes, but significant areas of overlap still exist. The two classes are best distinguished based on the hardness of their colour-colour diagrams (Belloni et al., 2000). Since both classes of observations show no structured X-ray variability patterns, and their IFoS distributions largely overlap, it should be noted that classification attempts based on features presented in this work could disagree with classification of Belloni et al. (2000). The proposed method is only able to capture time series patterns, therefore it would not be able to differentiate between observations which differ only in terms of energy spectra. Nevertheless, users of the method could choose to supplement SIFoS with additional features, dependent on their individual use case.

Another set of limitations involves data pre-processing. The re-binning of light curve data to 1 and 4 second resolution was a compromise between computational tractability and descriptiveness of GRS 1915+105 variability. Even though the shortest timescale of significant X-ray flux changes in this source is ca. 5 seconds (Nayakshin et al., 2000), binning inevitably led to a loss of some fast variability information. Furthermore, the choice of light curve segment length was influenced by computational constrains, but also informed by the previous work and knowledge of the time scales of light curve patterns in GRS 1915+105. There is a fairly large degree of tolerance for the choice of these parameters, but some knowledge of the time scales of interest was required.

I did not fine-tune the method to study any particular subset of variability classes within the source, however users interested in a more specialised study of a particular scientific problem could choose to analyse light curves binned to smaller time bins to improve the modelling of very narrow features. For example, Massaro et al. (2020) modelled the light curves of GRS 1915+105 using a system of ordinary differential equation, and successfully reproduced the sharp peaks present at the beginning of the κ class bursts. Reconstruction of light curve segments binned to 1 second resolution results in some smearing of such fine features (see Figure 2.4), so the study of this particular feature would benefit from higher time resolution. However, the light curves of GRS 1915+105 contain a significant amount of noise, and increasing the temporal resolution of light curves will result in the increased amount of noise in the training data set, which could affect the ability of the model to identify patterns within the data.

For cases where relevant time scales are difficult to predict, multiple fingerprint could be derived from data in a range of temporal resolutions, and subsequently concatenated to create a single feature vector. However, such an approach would increase the noise in feature vectors, so further work is required to test this.

My main assumption was that light curve segments exhibiting similar type of variability patterns and count rate distribution had similar values of SFoS and IFoS. Segments with more similar values of SIFoS were consequently separated by smaller distances within the SIFoS feature space. Therefore, it was expected that Gaussian clusters contained homogeneous subsets of light curve segments, which were more similar to each other than to segments found in other Gaussian clusters. Inspection of Gaussian cluster populations revealed that this assumption was justified to a large degree, however fitting a GMM with a larger number of Gaussian components would provide a greater precision of density estimation of data in the feature space, and hence reduce the risk of non-homogeneity of created clusters. On the other hand, such an approach would likely aggravate the issue of pattern degeneracy in the resulting set of light curve patterns, which would potentially lead to a more complex fingerprint and ultimately to a more noisy feature set.

It should be noted that GMM is not inherently a clustering algorithm but a density estimation algorithm. Following preliminary experiments with alternative clustering algorithms, I find the approach of merging Gaussian components of GMM to be the most straightforward method of clustering for this data set. However, it is possible that the issue of cluster degeneracy could be circumvented if a better-suited clustering method was found. This is an important area of future work.

Merging of clusters based on the Mahalanobis distance can help to reduce the number of degenerate features, but in this work I rely on the classified subset of data to find the optimal distance threshold. In cases where data classification is not possible, unsupervised methods of evaluating clustering performance could be used instead, for example Silhouette Coefficient, Calinski-Harabasz Index and Davies-Bouldin Index included in the SCIKIT-LEARN library (Pedregosa et al., 2011), however the computational cost of such an approach might be higher.

The timescales of X-ray variability probed by the method described in this chapter range from 1 second (resolution of the analysed light curves) to 128 seconds (size of a single light curve segment fed to the neural network). However, in the context of the fingerprint representation, the longest timescale could be several thousand seconds, because the fingerprints were generated for entire light curves. Timescales of structured variability found in the light curves of GRS 1915+105 range from ~20 s in case of the flares of the μ class, to ~1000 s in case of the cycles found in the α class.

The mass of $\sim 12.4 M_{\odot}$ (Reid et al., 2014) corresponds to the gravitational radius (R_g) of 18.3 km. It means that the timescale of 1 s corresponds to $1.6 \times 10^3 R_g$. GRS 1915+105 has a very large disc radius of $R_T \sim 10^{10}$m, which depends on the binary separation, because of the tidal truncation (Done et al., 2004). It means that light should travel from the black hole to the edge of the disc in ~33.4s.

Chapter 3

Light curve fingerprint method applied to IGR J17091−3624

The previous chapter introduced the "light curve fingerprints"; a standard, vector representation of light curves with variable length. It demonstrated how the fingerprints can be used for the analysis of X-ray variability of the Galactic low-mass black hole X-ray binary system GRS 1915+105. However, in addition to the analysis of individual sources, the light curve fingerprints could be a useful tool for the analysis of large data sets, containing observations of multiple sources, with various types of variability and intensity of signal. Since this methodology is based exclusively on the time series information, it should be possible to extend the analysis to other sources, after appropriate scaling of light curve intensity. In order to test the fingerprint representation on data of another source, I apply it to the RXTE observations of IGR J17091−3624, which also shows very complex patterns of flux variability, similar to those of GRS 1915+105. I investigate whether the method would be able to identify similar activity in the two sources, and hence test what adjustments need to be made in order to extend the use of fingerprint representation to other sources.

3.1 Can the fingerprints find similarity between IGR J17091−3624 and GRS 1915+105?

IGR J17091−3624 is a black hole LMXB whose X-ray variability has been found to show great similarity to that of GRS 1915+105 (Altamirano et al., 2011). The variability of both of the sources has been categorised into discrete systems of classes (Belloni et al., 2000; Court et al., 2017). I use the RXTE observations of IGR J17091−3624 from year 2011 (with observation ID 96420-01) to test the fingerprint method of light curve characterisation. Observations with ID numbers 96065-03 and 96103-01 are excluded because of contamination caused by GX 349+2, another variable X-ray source that was present in the field of view (Altamirano et al., 2011;

TABLE 3.1: The nine variability classes of IGR J17091−3624 introduced by Court et al. (2017), and the closest corresponding variability class in GRS 1915+105 (as introduced by Belloni et al. (2000)). Court et al. (2017) did not find analogies to the remaining GRS 1915+105 classes in IGR J17091−3624. Credit: Court et al. (2017)

IGR J17091−3624 Class	GRS 1915+105 Class
I	χ
II	ϕ
III	ν
IV	ρ
V	μ
VI	λ
VII	*None*
VIII	*None*
IX	γ

Rodriguez et al., 2011; Court et al., 2017). I aim to verify whether the method can be used to classify observations of this source, which shows variability that is similar to the one found in the training data set of GRS 1915+105 observations that was used to condition the method.

3.1.1 Classification based on variability shape and intensity information

In this first classification experiment, I follow the steps of data preparation and classification as described in Sections 2.2 to 2.4.3, treating the IGR J17091−3624 data as if they were new observations of GRS 1915+105. The only modification is the scaling of the IGR J17091−3624 data, which is applied to account for the significant difference in brightness between the two sources. I scale and offset the count rates across all the IGR J17091−3624 segments to make their mean and standard deviation identical with those of GRS 1915+105. In summary, I create the fingerprint representations of 215 IGR J17091−3624 observations by following these steps: re-binning of light curves to 1 second bins, segmentation (using the segment length of 128s, stride length of 10s), scaling to GRS 1915+105 intensity, encoding of segments using the LSTM-VAE, clustering of standardised latent variables and intensity features using the GMM that was fit to the GRS 1915+105 data, and finally merging the clusters using the Mahalanobis distance threshold of 3.34.

Once the fingerprints of IGR J17091−3624 observations are constructed, I proceed with the classification step. The ground truth class assignment of IGR J17091−3624 light curves is taken from Court et al. (2017). Court et al. (2017) created a set of nine variability classes, denoted by Roman numerals I - IX, which describe the different types of variability observed in the source. Court et al. (2017) also provided their classification for the 215 observations. Additionally, Court et al. (2017) matched the IGR J17091−3624 classes with GRS 1915+105 classes, based on the similarity of their spectral hardness evolution as a function of time and count rate, power spectral features and light curve morphology.

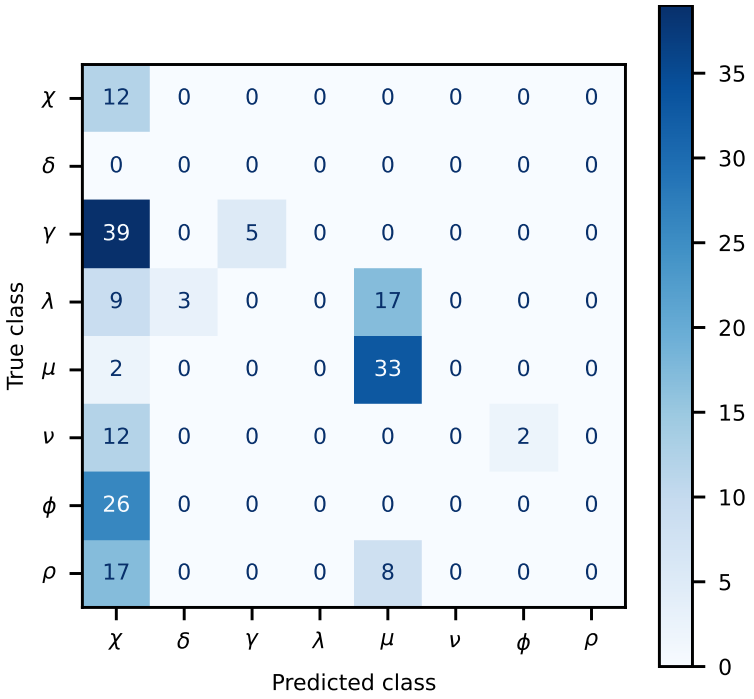


FIGURE 3.1: A confusion matrix showing classification results for the 185 IGR J17091−3624 observations with a matching GRS 1915+105 class, using the default fingerprint representation that contains both intensity and shape information. Weighted F1 and accuracy of this matrix are 0.195 and 0.270 respectively. The mean weighted F1 and accuracy performance scores across the one thousand initiations of the random forest classifier are 0.193 ± 0.024 and 0.268 ± 0.014 respectively.

Table 3.1 provides the matching variability classifications from the two sources. Based on this information, I assign the Belloni et al. system classification to each observation of IGR J17091−3624. Since classes VII and VIII do not have corresponding GRS 1915+105 classes, I exclude those from the classification experiment, which leaves 185 observations. I train a random forest classifier on the whole available set of 208 classified GRS 1915+105 fingerprints and use it to predict the classes of IGR J17091−3624 fingerprints. As mentioned in previous sections, random initiation of the random forest classifier is a significant cause of variance in the model performance, therefore training and testing is repeated one thousand times. Mean of weighted F1 and accuracy performance scores of those classifications are 0.193 ± 0.024 and 0.268 ± 0.014 respectively.

Figure 3.1 shows the classification results of an experiment with average performance scores in the form of a confusion matrix. It is clear that the classifier is biased towards the χ and μ classes; the recall (i.e. proportion of true positives out of the sum of true positives and false negatives) of χ is 1.0 and the precision (proportion of true positives out of all positives) is only 0.103. μ class has recall of 0.943 and precision of 0.569. The classifier showed some limited ability to differentiate the γ class, with recall of 0.128. The classification of ρ observations is one of the most striking results; ρ is the most characteristic class of variability in GRS 1915+105, due to its recurrent flares. However, the classifier assigns most of the ρ observations to the χ class, and some to the μ class.

The fact that some ρ class observations are classified as μ could be caused by the shorter period of heartbeats in IGR J17091−3624. Some μ observations in GRS 1915+105 show recurrent flares with a period of ∼20 seconds, which is similar to the heartbeat of IGR J17091−3624. However, the large number of χ classifications suggests that the classifier was not able to utilise the information about the shape of the variability patterns in the light curves. Similarly to the ρ class of GRS 1915+105, the heartbeat class of IGR J17091−3624 (class IV) shows very characteristic, recurrent flares. They tend to be less regular and have a shorter period than those of GRS 1915+105, but nevertheless they differ significantly from the χ-like variability, which is random and has no structure. The inability of the classifier to differentiate these two classes indicates that the SFoS were either not informative enough or underutilised.

Even though the light curve segments were scaled to have the same mean as the GRS 1915+105 segments, this did not guarantee that the ρ class observations of the two sources would show similar intensities. In order to achieve this, the observations of different classes would need to be scaled separately, but this is not viable, because it would require the prior knowledge of light curve classifications. In order to prevent the IFoS from causing the observations of the same classes to become misaligned in the feature space, and to test whether the SFoS of IGR J17091−3624 contain enough information to match its light curves with the most similar classes of GRS 1915+105, I perform a classification experiment using SFoS alone.

3.1.2 Classification based on variability shape information alone

In order to construct the fingerprints of IGR J17091−3624 observations based on SFoS alone, I follow the same steps as in the previous section, but this time I fit a GMM model to the SFoS of the training set of GRS 1915+105 light curves. Instead of performing a grid search for the optimal number of Gaussian components, I fit one GMM with 500 components, which is well above the expected number of components that would produce a minimum BIC value. This is an alternative approach that reduces computational time, and is valid thanks to the cluster merging step applied later.

This model fitting step is followed by the same procedure as described earlier, i.e. I first find the assignment of each light curve segments to a Gaussian component. Then, I search for the optimal Mahalanobis distance threshold in a grid search over the hyperparameters given in Table 2.4, using the training and validation subsets of GRS 1915+105 data. When the optimal distance is identified, I use it to construct the fingerprints of the observations of both sources. Fingerprints of GRS 1915+105 sources are constructed in order to verify whether the modified representation still allows for reasonable classification performance, and to use it as a benchmark. I classify the test set of GRS 1915+105 observations to find this benchmark performance score.

Mean of weighted F1 and accuracy performance scores of those classifications are 0.810 ± 0.035 and 0.842 ± 0.025 respectively. Figure 3.2 shows the classification results with the lowest performance scores out of the set of 1,000 testing trials. The weighted F1 and accuracy of this

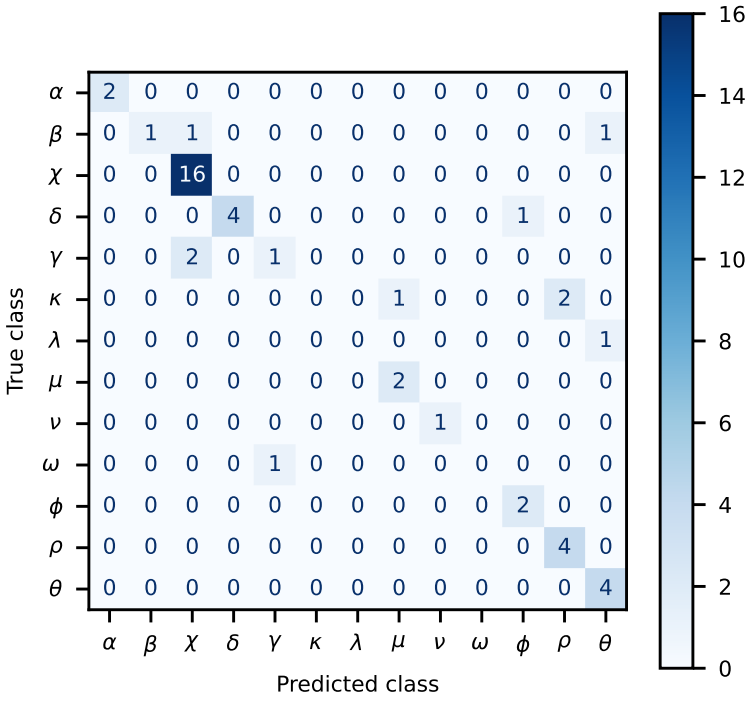


FIGURE 3.2: Confusion matrix showing classification results for the 47 testing set observations of GRS 1915+105, using the fingerprint representation that contains shape information (SFoS features) only. The matrix shows results for the lowest weighted F1 score out of the one thousand random initiations of the classifier. Weighted F1 and accuracy of this initiation are 0.731 and 0.787, respectively. The mean weighted F1 and accuracy performance scores across the 1000 initiations are 0.810 ± 0.035 and 0.842 ± 0.025, respectively.

initiation are 0.731 and 0.787 respectively. The results show that the SFoS provide the classifier with enough information to identify the observations of classes χ, ρ and θ with maximum recall of 1.0, whilst δ has a high recall of 0.8. Classes α, μ, ν and ϕ also show 1.0 recall, but for a small test sample size of only $1-2$ observations.

This experiment confirms that the shape information alone might not be enough to differentiate class γ from χ even for the GRS 1915+105 observations, which the method is conditioned on. However, the classifier shows no signs of confusion between classes χ, μ, ϕ and ρ, which indicates that observations showing characteristic patterns of variability should be classified correctly even if observed at different intensities. Therefore, if patterns characteristic to those classes were observed in IGR J17091−3624, then the method should be able to classify them correctly based on their SFoS only.

After having verified that the new fingerprints can provide a good classification performance on the GRS 1915+105 data, I perform the classification of 185 observations of IGR J17091−3624. The mean weighted F1 and accuracy performance scores of the one thousand initiations are 0.204 ± 0.011 and 0.257 ± 0.009. The confusion matrix of an example initiation are shown in Figure 3.3. The majority of observations are again classified as classes χ and μ, but this time the classifier also shows bias towards the class γ, with many λ, μ and ρ observations labelled as γ, despite the fact that the class still has a low recall of 0.16. A majority of observations

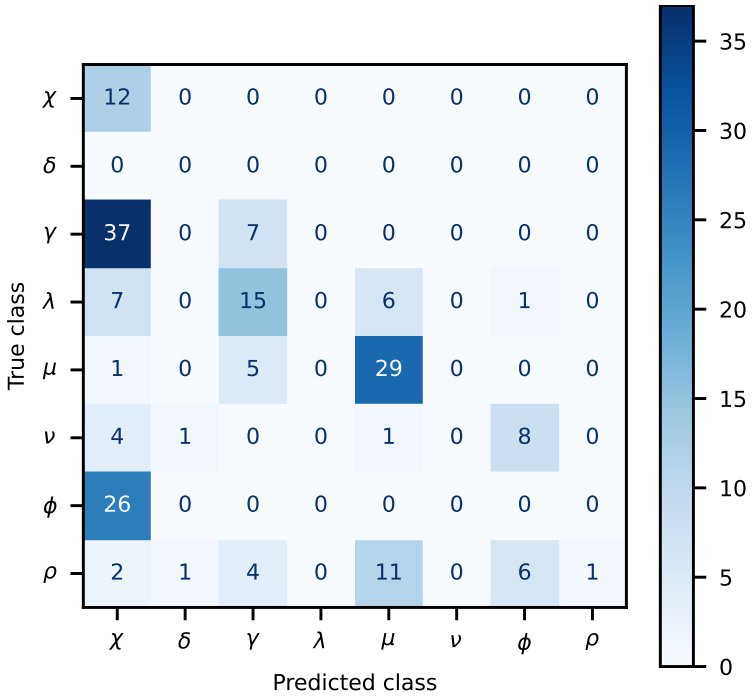


FIGURE 3.3: A confusion matrix showing classification results for the 185 IGR J17091−3624 observations with a matching GRS 1915+105 class, using the fingerprint representation that contains shape information only. Weighted F1 and accuracy of this initiation are 0.204 and 0.265 respectively. The mean weighted F1 and accuracy performance scores across the one thousand initiations of the random forest classifier are 0.204 ± 0.011 and 0.257 ± 0.009 respectively

of class ρ are now assigned to class μ, with the remaining observations distributed across all the other classes. The classification of IGR J17091−3624 observations using exclusively SFoS information shows that the lack of IFoS information makes very little difference to the overall performance of the classifier.

The disagreement between Court et al. (2017) classifications and those predicted with fingerprints could be caused by the LSTM-VAE model finding significant differences between the light curves of GRS 1915+105 and IGR J17091−3624. In such a case, SFoS generated from the light curves of the two sources would differ significantly. Examination of the light curves of the two sources shows that they indeed differ in several aspects. For instance, the flares in ρ class light curves of IGR J17091−3624 are less regular in shape than those of GRS 1915+105. The period of flares is also significantly shorter; the ρ flares in GRS 1915+105 have periods ranging between ∼40 – 150 s, while μ flares range between 10 – 100 s (Belloni et al., 2000). The periods in corresponding classes of IGR J17091−3624 range between ∼20 – 45 s and 10 – 30 s respectively. Additionally, the count rate levels of IGR J17091−3624 are 10 – 50 times lower than those of GRS 1915+105 (Altamirano et al., 2011), resulting in a smaller signal-to-noise ratio and a much greater amount of noise in the scaled light curves. In order to investigate if the unexpected classification is caused by the differences between the variability of the sources, I will first examine whether the LSTM-VAE network can successfully encode and reconstruct the variability of IGR J17091−3624. Once that is verified, I will examine how the encoding (SFoS) of the two sources differ.

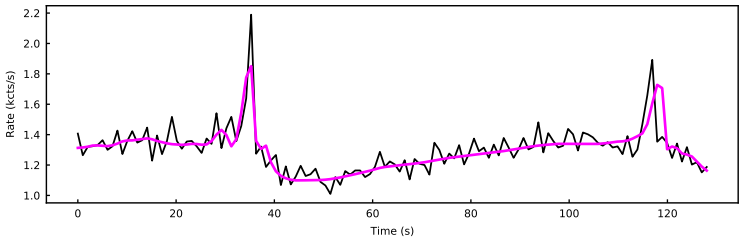
3.1.3 Encoding the variability of IGR J17091−3624

Figure 3.4 shows the LSTM-VAE reconstruction of some example light curve segments of IGR J17091−3624 which show variability of classes III (equivalent to ν), IV (equivalent to ρ) and V (equivalent to μ), all of which show characteristic quasi-periodic flares similar to those found in GRS 1915+105. LSTM-VAE is able to reconstruct the majority of flaring behaviour in IGR J17091−3624 well. Figure 3.4 segments (b), (d) and (e) show examples of flares with various periods and morphologies, whose reconstruction was successful most of the time. Segment (e) shows a reconstruction where the LSTM-VAE attempts to fit a dip in the rising phase of each flare, which could be an attempt of fitting a double-peak that is characteristic of ρ flares in GRS 1915+105 (Neilsen et al., 2011), although it is also observed in other classes; see for example class μ in Figure 2.4. Something about the morphology of flares in (e) causes the model to associate it with the double-peaked profile, in spite of the fact that it systematically underestimates the count rates in the modelled dips. Segment (f) shows that the model struggles to reconstruct the flares when they are less regular. The reconstruction of IGR J17091−3624 light curve segments shows that the LSTM-VAE is capable of capturing the variability structure of a significant number of observations showing quasi-periodic flaring. Therefore, disagreement in the classification of GRS 1915+105 and IGR J17091−3624 observations cannot be explained by a failure of the neural network to capture the variability of IGR J17091−3624.

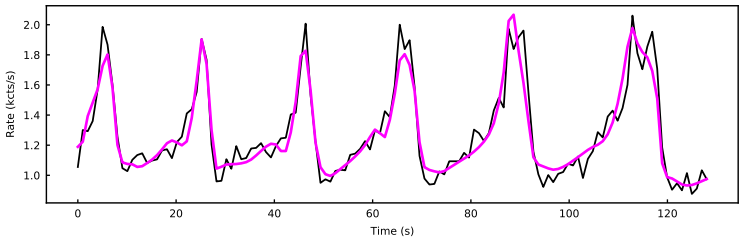
In order to shed some light on the reason the SFoS do not allow for the differentiation of μ from ρ in IGR J17091−3624, let us now examine how the variability of input data affects their position in the latent space of the LSTM-VAE. Even though the VAE architecture is meant to prevent this type of behaviour, it is still hypothetically possible for an autoencoder network to map similar data to disjoint locations of the latent space, causing substantial differences in SFoS values, in spite of a seemingly small difference in the appearance of light curve segments. It is also possible that segments of a light curve could span multiple regions of the latent space, and collectively look like a different class in the context of the fingerprint.

Figure 3.5 shows the UMAP projection of the SFoS of both the GRS 1915+105 and IGR J17091−3624 segments. GRS 1915+105 segments showing the most characteristic variability classes are coloured. It is apparent that all classes show significant scatter across the latent space, but the bulk of segments of these characteristic classes aggregates in discrete sections of the latent data manifold; class ρ, which shows the most regular, recurrent flares, occupies the two outer folds on top and to the right of the data cloud. The central section is occupied by classes with less regular flaring behaviour, μ, ν, and κ, with a significant portion of ν occupying the outer folds as well. The circular structure in the bottom-left is primarily composed of class χ, which shows no structured variability.

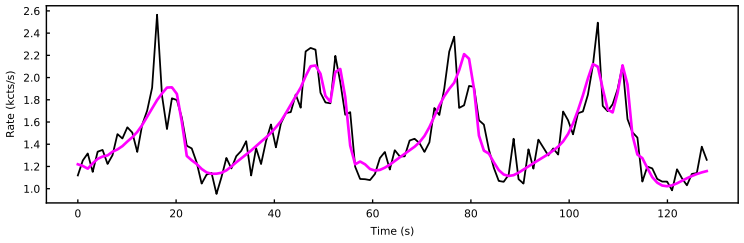
Now, keeping in mind the relationship between the GRS 1915+105 classes (and their variability patterns), and the latent space regions, I examine the distribution of class IV (ρ) segments of IGR J17091−3624 in this space, in order to verify if they occupy similar regions of the SFoS space as the GRS 1915+105 segments of the same class. Observations of class IV can be roughly



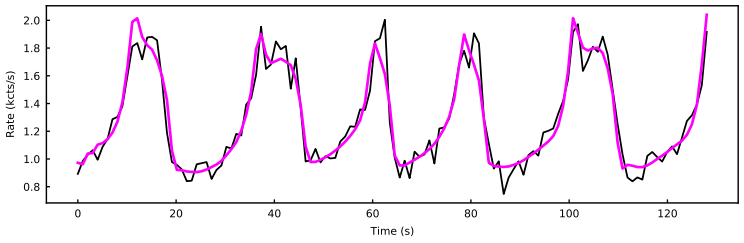
(a) Observation ID: 96420-01-04-01, Class: III



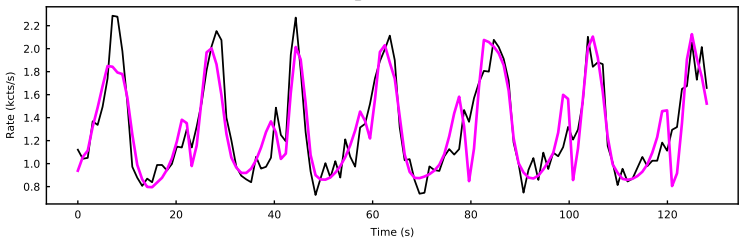
(b) Observation ID: xp96420010601, Class: IV



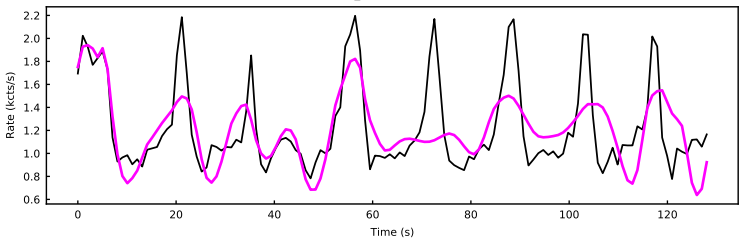
(c) Observation ID: xp96420010400, Class: IV



(d) Observation ID: xp96420010700, Class: V



(e) Observation ID: xp96420012504, Class: V



(f) Observation ID: xp96420011804, Class: V

FIGURE 3.4: Examples of reconstruction of IGR J17091−3624 light curve segments for observations classified as IV (ρ) and V (μ). The input data is shown in black and the reconstruction in magenta.

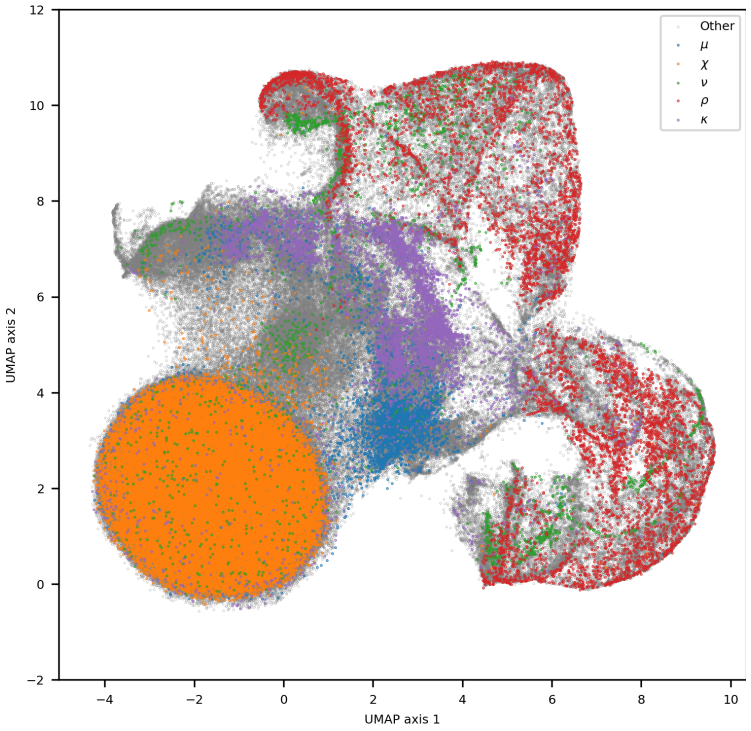


FIGURE 3.5: UMAP projection of SFoS from segments of GRS 1915+105 and IGR J17091−3624. Classification of the GRS 1915+105 segments is indicated with colour for some characteristic classes.

grouped into four types with respect to their segment distribution in the SFoS space. Figure 3.6 shows the segment distribution of four example observations:

- (3.6a) The light curve segments of this observation show mostly regular flares with a period of ∼27 s. The morphology of flares does not change significantly within a segment, and many show the characteristic ρ-like double peak (see an example in Figure 3.4b), with few segments showing abrupt changes in the flare intensity, duration, and shape. Its segments are clustered in the μ-dominated, lower-central region.

- (3.6b) Segments of this observation show clear quasi-periodic oscillation with a period of ∼30 s, with some segments showing flares that can change morphology or stop abruptly. Segments found in the region associated with χ-like variability can show very irregular behaviour, but some also contain intervals with very well-defined flares, which do not last for the duration of a whole segment. Its segments are scattered across multiple regions, including those associated with χ, μ and ρ-like variability.

- (3.6c) Light curve segments of this example observation show fairly regular flares with a period of ∼32 s, a slow rising phase, and one or multiple sharp peaks in the maximum intensity phase (see an example in Figure 3.4c). Its segments are clustered around the intersection of the outer folds and the inner region of the manifold, bridging the regions associated with the μ and ρ variability.

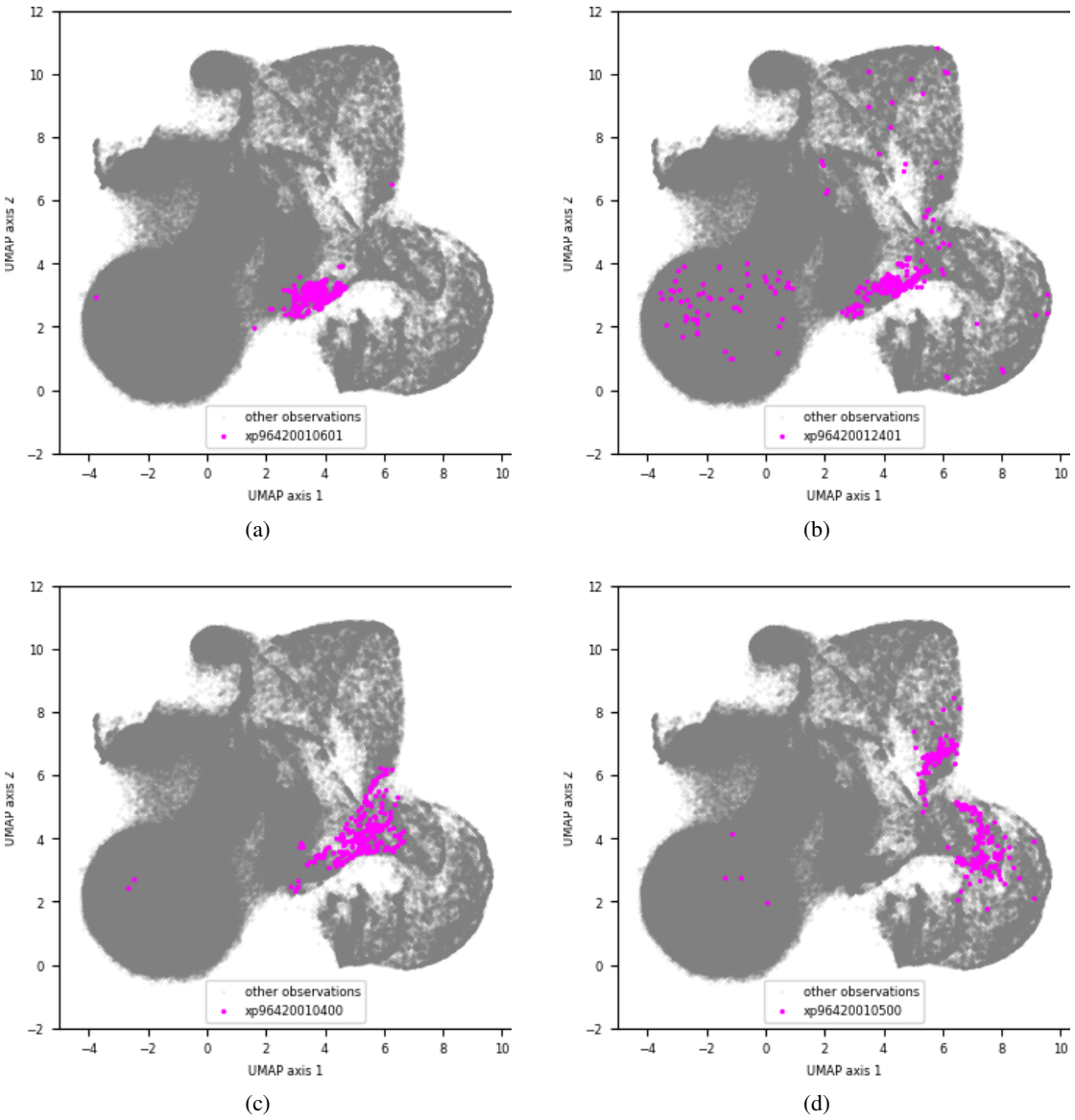


FIGURE 3.6: UMAP of SFoS of four example IGR J17091−3624 class IV observations

- (3.6d) Segments of this observation show clear quasi-periodic oscillation with a ∼43 s period, even though the morphology of the flares is quite irregular. The flares tend to have a slowly rising profile with a sharp peak. Its segments are in the ρ-dominated outer folds, but close to the joining point of the two folds.

The position of light curves in the SFoS space shown in Figure 3.6 reveals that the segments move from the μ-dominated region out to the ρ-dominated region as the period of flaring increases. In an ideal scenario, the LSTM-VAE would map the frequency of signal to one of the latent variables, and thus completely disentangle this parameter of the signal from others, like phase or morphology. In case of this network, the frequency of signal is not controlled by a single variable, but multidimensional regions of the latent space seem to correlate with the frequency of signal. This observation points to a possible explanation for the fact that most of the class IV (ρ) observations of IGR J17091−3624 were classified as μ, which could be due to the flaring

frequency. There are at least two possible approaches to testing this. One way involves generating synthetic light curves based on the IGR J17091−3624 data and increasing their period to track how their position in the SFoS space changes. Alternatively, the latent space could be probed using the LSTM-VAE decoder, which could be used to generate light curve segments from arbitrary SFoS values, and hence sample light curve segments from the regions which seem to be associated with changing frequency of signal. The first method is preferred, because it gives more insight into the workings of the encoding block of the LSTM-VAE, which is more relevant to the feature extraction process and light curve classification. In the next section I investigate the effect of changing signal frequency on the SFoS values and the resulting classification.

3.1.4 Effect of changing flare period on light curve classification

In order to test whether the period of the flare could be a major factor that causes the neural network to associate the flaring light curves of IGR J17091−3624 with other classes than the ρ class of GRS 1915+105, I synthesise new light curves from the data of IGR J17091−3624, artificially increase the period of the flares in the light curves, and then track their position in the SFoS space. The synthesis method finds the time bins with count rates at the base level (outside the flare) in order to elongate those base level sections of the light curve, and to minimise changes to the morphology of the flares themselves. In order to achieve the desired period of the flares in the synthesised light curves, I insert additional bins in random positions of the light curve, where the number of inserted bins is equal to

$$b_{insert} = b_{old}\left(\frac{p_{new}}{p_{old}}\right) - b_{old} \tag{3.1}$$

where b_{insert} is the number of bins to be inserted, b_{old} is the number of bins in the original light curve, p_{new} is the new, desired period, and p_{old} is the period in the original light curve. The probability of insertion is weighted using the probability density function of the count rates to the third power, so that the insertion is most likely to happen in the positions where the count rate is closest to the mode count rate value (most popular value in the distribution). I use the log-normal distribution to approximate the distribution of count rates in the light curve. The probability density function of a log-normally distributed random variable x is

$$f(x) = \left(\frac{1}{\sigma(x-\theta)\sqrt{2\pi}}\exp\left(-\frac{\ln^2\left(\frac{x-\theta}{m}\right)}{2\sigma^2}\right)\right) \tag{3.2}$$

for $x > \theta; \sigma, m > 0$, where θ is the location parameter, σ is the shape parameter, and m is the scale parameter (Heckert and Filliben, 2003). The probability density function is parameterised separately for each light curve, so that θ is set to the mean of the natural logarithm of the count rate, σ is set to the standard deviation of the natural logarithm of the count rate, and m is set to the exponential of θ. I find the log-normal likelihood of each observed count rate value in

the 0.125 second bins, and then cube it, in order to make the values farther from the mode of the distribution even less likely. Then I select b_{insert} bins to be replicated, through random selection of light curve bins with replacement, where the probability of selecting each bin is set to the cubed log-normal likelihood of that bin divided by the sum of log-normal likelihood values across all bins.

New bins are inserted in the selected positions, and the count rate values of the new bins are set to the moving average value at that selected position, where the moving average is calculated using a window with the size of 11 bins. The moving average is used instead of the original count rate values in order to minimise the noise in the synthetic light curve at this stage; initially, I want to simulate data that follows the general trend of the light curve. Subsequently, I add random noise to the inserted count rate bins, where the magnitude of the noise is a random value drawn from the normal distribution with the mean of zero, and the standard deviation equal to that of the count rate distribution in the original light curve. The noise values are then adjusted to match the fractional root-mean-square (RMS) variability amplitude (F_{var}) of the original light curve. The F_{var} (Vaughan et al., 2003) is defined as

$$F_{var} = \sqrt{\frac{S^2 - \overline{\sigma_{err}^2}}{\bar{x}^2}} \tag{3.3}$$

where S^2 is the variance of light curve count rates, $\overline{\sigma_{err}^2}$ is the mean square measurement error, and $\bar{x}$ is the mean count rate. F_{var} is the square root of the rate-normalised excess variance,

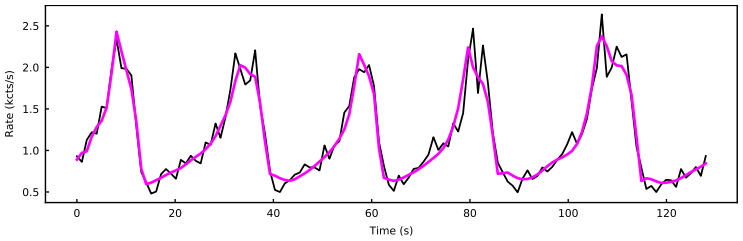
$$\sigma_{XS}^2 = S^2 - \overline{\sigma_{err}^2} \tag{3.4}$$

which is the variance with the subtracted contribution of Poisson measurement errors. I use F_{var} as a measure of variance in the light curves, and I scale the magnitude of noise added to the synthesised light curve in order to minimise the difference between the F_{var} of the observed light curve and the synthesised light curve.

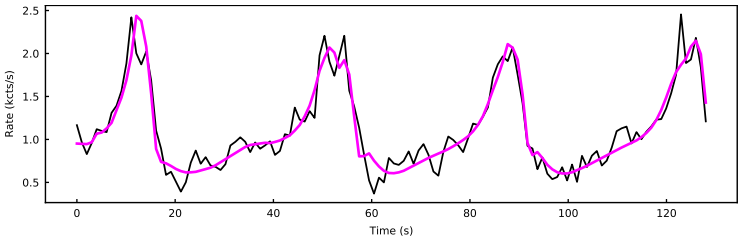
A scaling factor is found for every synthesised light curve using the L-BFGS-B[1] (Zhu et al., 1997) algorithm implemented in the Python optimisation/minimisation module of SciPy package (Virtanen et al., 2020). The lower and upper bounds for the scaling factor value are set to 0 and 1 respectively, and the initial guess is set to 0.5. The algorithm has two termination conditions, where one tests for the change in the value of the objective function over one iteration, and the second tests for the magnitude of the projected gradient, and they are both set to 0.01. Once the noise multiplication factor is found, I use it to scale the Gaussian noise added to the new bins. Finally, I re-bin the synthetic light curves to 1 s temporal resolution.

I synthesise light curves with elongation factors of 1.0, 1.2, 1.5, 1.8, 2.0, 3.0, 4.0, and 5.0 using each one of the 25 IGR J17091−3624 heartbeat light curves as the starting point. Figure 3.7 shows

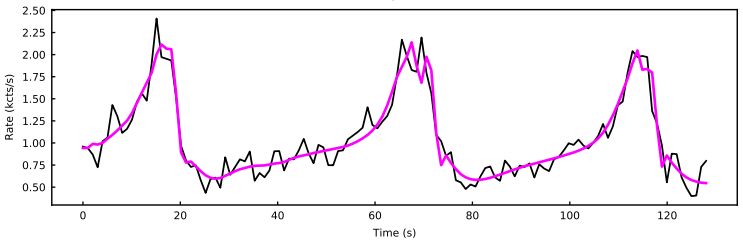
[1]L-BFGS-B is the limited memory version of the Broyden–Fletcher–Goldfarb–Shanno minimisation algorithm that allows for the setting of bounds on the parameter space. See Byrdt et al. (1995) for more details about the algorithm.



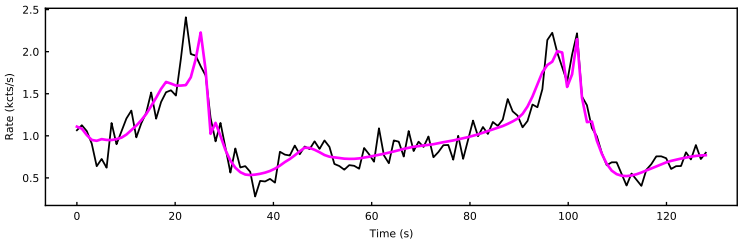
(a) ObsID: 96420-01-06-01, Class: IV



(b) Period increased by the factor of 1.5



(c) Period increased by the factor of 2.0

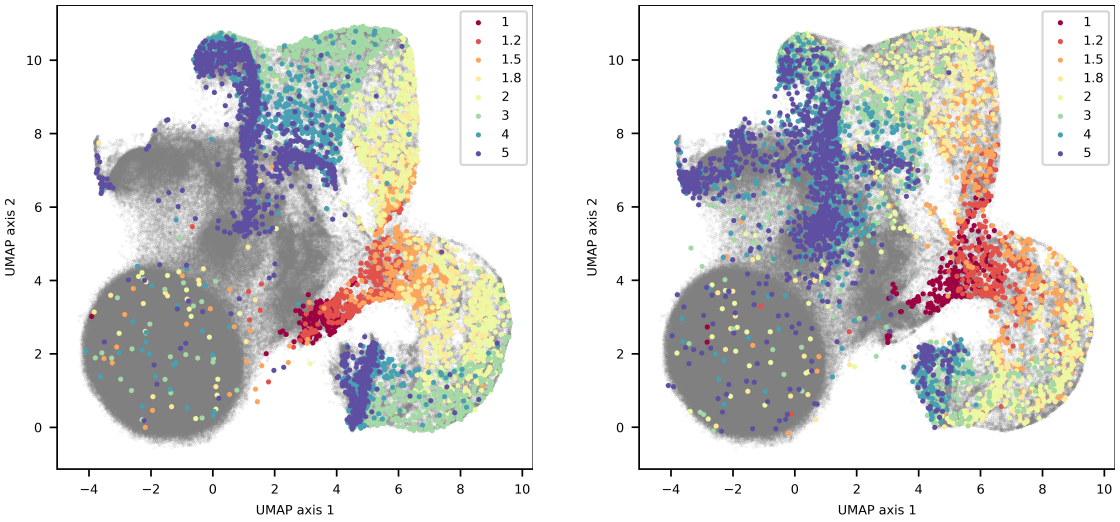


(d) Period increased by the factor of 3.0

FIGURE 3.7: First 128 s of IGR J17091−3624 observation 96420-01-06-01 classified as IV (ρ) and synthetic light curves with increased period generated from that observations. LSTM-VAE reconstruction is included; the input data is shown in black and the reconstruction in magenta.

examples of synthetic data: Subfigure 3.7a shows an original 128 s segment from observation 96420-01-06-01, while Subfigures 3.7b to 3.7d show segments of synthetic light curves generated from that observation, which are made with elongation factors 1.5, 2.0 and 3.0. It is clear that changes to the morphology of flares are not completely avoided in the synthesis process, but the periodicity of the signal is retained, and the peaks are still very well-defined.

Figure 3.8 shows the UMAP projection of SFoS of the light curves synthesised from observations 96420-01-06-01 and 96420-01-04-00. Segments of light curves with increasing flare period are marked with different colours. The resulting figures confirm the previous inferences about the position of ρ and μ segments of GRS 1915+105 light curves, i.e. light curve segments that show quasi-periodic flares are mapped to specific regions of the SFoS space, where the bottom central region of the manifold is associated with the shortest periodic signals found in the GRS



(a) ObsID: 96420-01-06-01 (b) ObsID: 96420-01-04-00

FIGURE 3.8: UMAP of the SFoS variables of two IGR J17091−3624 observations of class IV (ρ), plotted together with the synthetic light curves generated based on those observations, were the period of flaring in the synthetic light curves was increased. The original periods of observations 96420-01-06-01 and 96420-01-04-00 are 27.27 s and 31.65 s respectively. Different colours indicate different factor by which the period of the light curve is increased compared to the original. Grey data points in the background are SFoS of all GRS 1915+105 and IGR J17091−3624 segments.

1915+105 μ class, and the segments with longer periods are associated with the outer folds to the right and top of the manifold, with progressively longer periods mapped farther out from the centre. UMAP of synthetic light curves generated from other observations show similar trends.

The segments showing the longest period flares are mapped to the ends of the outer folds of the SFoS manifold, and beyond a certain period length, an increasing proportion of segments gets mapped to the central region of the manifold. This period at which the transition happens is most likely dictated by the light curve segment length used in the LSTM-VAE. In fact, Figure 3.8a shows this transition; the segments of light curve with the elongation factor of 4 (resulting in a period of ∼109 s) are mostly positioned in the outer folds of the SFoS manifold, and the segments with elongation factor of 5 (which results in a period of ∼136 s) are clustered at the terminal ends of the outer folds and in the central region as well.

The position of synthetic light curve segments in the SFoS space seems to be a very good indicator of the light curve's variability pattern, and in case of segments showing recurrent flares, their relative period of the flare. It seems that the segments with shorter period are associated with the μ class region of the manifold, and the increasing period causes them to move to the regions associated with the ρ class. Therefore, the fact that the heartbeat observations of IGR J17091−3624 were rarely classified as ρ in the previous experiments seems to be caused by the fast periods of quasi-periodic flares. In order to confirm this, I again use a random forest classifier trained on all 208 GRS 1915+105 observations to predict the class of synthetic IGR J17091−3624 light curves. I perform classification 1000 times to account for the effect of the random initiation of the algorithm. Table 3.2 shows the classification results, and it lists the

Observation ID	Period (s)	Classification (probability) for light curves with changing period elongation factor							
		1.0	1.2	1.5	1.8	2.0	3.0	4.0	5.0
96420010400	31.65	$\mu(0.7)\delta(0.2)$ $\rho(0.1)$	δ	ϕ	ϕ	ϕ	$\phi(0.6)\alpha(0.4)$	α	δ
96420010402	38.32	$\mu(0.8)\rho(0.2)$	$\delta(0.9)\phi(0.1)$	ϕ	ϕ	ϕ	$\alpha(0.8)\phi(0.2)$	δ	$\delta(0.9)\phi(0.1)$
96420010500	43.46	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
964200105000	40.88	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
96420010501	37.64	$\phi(0.7)\delta(0.3)$	$\rho(0.8)\phi(0.2)$	$\phi(0.7)\delta(0.3)$	ϕ	ϕ	α	α	δ
96420010503	39.40	$\phi(0.8)\rho(0.1)$ $\delta(0.1)$	ϕ	ϕ	ϕ	ϕ	ϕ	$\phi(0.8)\delta(0.2)$	$\phi(0.6)\delta(0.4)$
96420010504	36.21	$\delta(0.9)\phi(0.1)$	$\rho(0.8)\chi(0.1)$	$\delta(0.5)\phi(0.3)$ $\rho(0.2)$	ϕ	ϕ	$\alpha(0.5)\phi(0.5)$	α	α
96420010600	27.59	μ	μ	ρ	ρ	$\nu(0.4)\phi(0.4)$ $\delta(0.1)$	ϕ	α	α
96420010601	27.27	$\rho(0.6)\mu(0.4)$	μ	μ	ρ	ρ	ρ	α	α
96420010602	26.58	μ	$\mu(0.8)\rho(0.2)$	μ	$\rho(0.7)\mu(0.3)$	ρ	$\alpha(0.6)\phi(0.2)$ $\nu(0.2)$	α	α
96420011305	23.34	γ	$\mu(0.8)\gamma(0.2)$	$\phi(0.9)\mu(0.1)$	$\phi(0.9)\chi(0.1)$	ϕ	ϕ	ϕ	ϕ
96420011400	23.73	$\gamma(0.6)\chi(0.3)$	$\mu(0.6)\chi(0.2)$ $\gamma(0.2)$	$\chi(0.8)\phi(0.2)$	$\gamma(0.5)\chi(0.5)$	γ	ϕ	ϕ	ϕ
96420011401	25.73	$\chi(0.9)\gamma(0.1)$	μ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
96420011402	25.49	μ	$\mu(0.9)\phi(0.1)$	ϕ	$\phi(0.9)$	ϕ	ϕ	$\phi(0.5)\alpha(0.5)$	α
96420011403	27.02	$\gamma(0.8)\mu(0.2)$	μ	μ	μ	$\phi(0.4)\gamma(0.3)$ $\mu(0.3)$	ϕ	ϕ	α
96420011500	40.92	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	$\delta(0.9)\phi(0.1)$
96420011501	33.80	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
96420011504	25.51	$\mu(0.5)\gamma(0.5)$	μ	ϕ	ϕ	ϕ	ϕ	$\phi(0.9)\alpha(0.1)$	$\phi(0.7)\alpha(0.3)$
96420011505	25.79	$\mu(0.8)\gamma(0.2)$	$\mu(0.9)$	ϕ	$\phi(0.9)\chi(0.1)$	ϕ	ϕ	ϕ	$\alpha(0.8)\phi(0.2)$
96420011600	28.57	χ	ϕ	ϕ	$\chi(0.9)\phi(0.1)$	ϕ	ϕ	ϕ	ϕ
96420011801	33.74	$\gamma(0.5)\chi(0.5)$	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
96420012306	30.31	$\chi(0.9)\mu(0.1)$	ϕ	χ	ϕ	ϕ	ϕ	ϕ	ϕ
96420012400	29.61	$\chi(0.8)\mu(0.2)$	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
96420012401	29.71	χ	$\phi(0.8)\chi(0.2)$	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
96420012402	26.95	μ	μ	$\phi(0.5)\mu(0.5)$	ϕ	ϕ	ϕ	ϕ	$\phi(0.9)\alpha(0.1)$

TABLE 3.2: Classification of the synthetic heartbeat observations of IGR J17091−3624 with the increased flare period. The classification with random forest classifier was repeated one thousand times, and whenever the proportion of predictions of a particular class was lower than 95%, multiple classes are listed with their proportion of predictions.

proportion of times that different classes are predicted for a given light curve, unless one class exceeds 95% of all classifications, and then only that one class is shown. Surprisingly, the effect of increasing flare period had a very limited effect on making the heartbeat IGR J17091−3624 light curves more similar to GRS 1915+105 ρ class. Out of the 25 observations, the only ones which are more likely to be classified as ρ following the increase in flare period are 96420-01-05-01, 96420-01-05-04, 96420-01-06-00, 96420-01-06-01 and 96420-01-06-02. In case of 96420-01-04-00, 96420-01-06-02 and 96420-01-05-03, the classification of ρ actually became less likely. A large proportion of the light curves is likely to be classified as ϕ.

3.2 Direction of further development

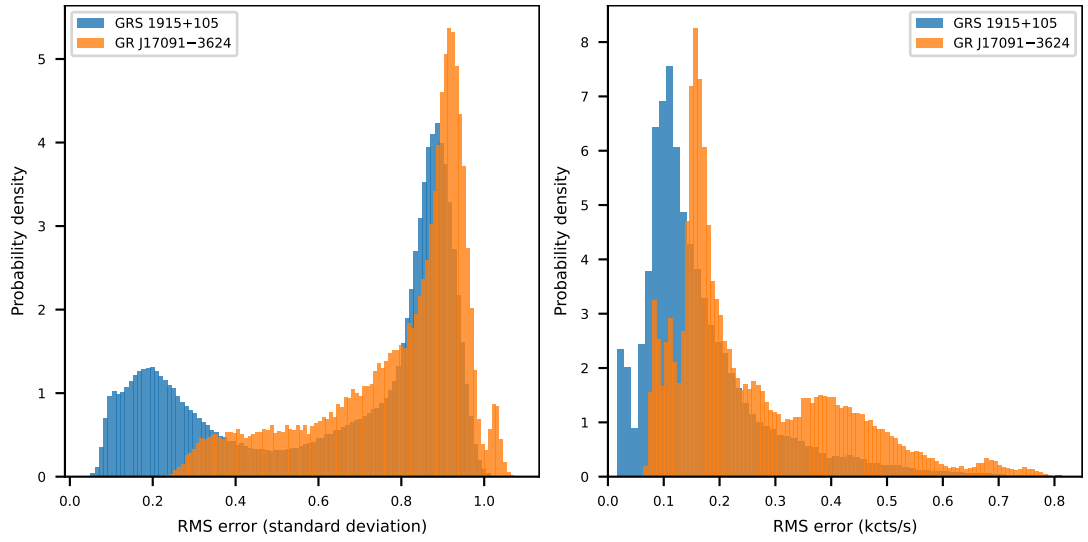
Classification experiments involving the observations of IGR J17091−3624 show that the fingerprint method of light curve characterisation, in its current state of development, does not find similarity between the variability of IGR J17091−3624 and GRS 1915+105. The desired outcome of the classification should draw links between the similar variability classes of the two sources, mimicking the outcome of a manual analysis of the data by an astronomer, just like in the work of Court et al. (2017).

Exclusion of the intensity information from the fingerprint representation is a natural step in the development of the method, since different sources analysed by the method will show variable intensity levels. Intensity features provide useful information during the analysis of the variability of a single source, but the natural conclusion that transpires from the experiments shown in this chapter is that the characterisation of variability of different sources should be based on the outputs from the LSTM-VAE alone.

Projections of data in the LSTM-VAE latent space show that the network can successfully encode information about the frequency of quasi-periodic behaviour. However, further investigation is needed to probe the reasons why fingerprints of IGR J17091−3624 light curves showing such behaviour are not classified as the ρ class, which would be the expected outcome for the "heartbeat" type of observations.

3.2.1 Quantifying reconstruction accuracy of light curve segments using LSTM-VAE

In order to quantify the reconstruction accuracy of the IGR J17091−3624 light curve segments, and compare it to the reconstruction of GRS 1915+105 segments from the test subset, I calculate the RMS difference (error) between every light curve segment and their LSTM-VAE reconstruction. Figure 3.9a shows the histograms of RMS error for all light curve segments of both sources in the units of standard deviation (like the LSTM-VAE input and output), whereas Figure 3.9b shows the same errors that have been scaled back to the original count rate units. The median RMS error of GRS 1915+105 and IGR J17091−3624 segments are 0.79 σ (133.1 counts/s) and



(a) Standardised reconstruction error

(b) Reconstruction error

FIGURE 3.9: RMS errors calculated between the light curve segments and their LSTM-VAE reconstructions. The RMS errors are calculated for (a) standardised light curve segments, and (b) segments scaled to the original count rate units. Only the GRS 1915+105 light curve segments from the test subset of data are included.

0.86 σ (200.8 counts/s) respectively. The probability density functions of RMS errors of the two sources mostly overlap, however, the average RMS error of GRS 1915+105 segments are significantly smaller than those of IGR J17091−3624. The GRS 1915+105 segments also have a significantly lower minimum RMS error of 0.05 σ (17.0 counts/s), compared to 0.19 σ (64.3 counts/s). However, GRS 1915+105 segments produce several outliers with high RMS errors in the count rate space; the maximum is 1260.0 (NB: the domain of Figure 3.9b is set to focus on the visible data, whilst truncating the high-error outliers). The maxima of the two histograms shown in Figure 3.9a are virtually identical, but the IGR J17091−3624 histogram shows a high-error spike.

The fact that the RMS errors of IGR J17091−3624 segments are significantly higher than those of GRS 1915+105 indicates that the LSTM-VAE is not able to reconstruct the light curves of the former source as well as the latter. A slight reduction in the accuracy of the reconstruction of IGR J17091−3624 light curve segments is expected, because the training set of data provided to the neural network did not contain any light curves of this source. The significant difference in the minimum RMS error between the sources is likely caused by the fact that the variability patterns found in the light curves of IGR J17091−3624 are more irregular and noisy.

Chapter 4

Searching for periodicities in the optical photometry of Be X-ray binaries

This chapter describes a method of automatic detection and characterisation of periodic signals of the type that can be found in the optical light curves of Be X-ray binaries (BeXRBs). The method consists of a pipeline of several algorithms which process the output of the weighted wavelet Z-transform (WWZ) (Foster, 1996). I apply the method to the I-band photometry of a sample of BeXRBs observed in the Small Magellanic Cloud as part of the Optical Gravitational Lensing Experiment (OGLE), and discuss what the results can reveal about the properties of these sources. Future photometric surveys will require automated methods of periodicity detection in order to identify different types of periodic behaviour in large data sets. The aim of the work described in this chapter is to develop a proof of concept pipeline that automatically characterises periodic signals which can be transient or change over time. The application of this pipeline to the sample of BeXRB light curves can contribute to the study of binary orbital behaviour and non-radial pulsations observed in those sources, and probe their relation to the physics of decretion discs in the Be stars.

This chapter contains the following sections: Section 4.1 introduces the X-ray Variables OGLE Monitoring System, where the analysed data sample was sourced from. It also describes the motivation for the analysis of this data, and the types of variability expected from BeXRB systems. Section 4.2 provides details about the data preprocessing. Section 4.3 describes the development of a new periodicity detection and characterisation pipeline that uses a new implementation of the weighted wavelet Z-transform. It also describes how the parameters of the pipeline are tuned using synthetic data. Section 4.4 presents the results of application of the pipeline to the sample of OGLE light curves of BeXRBs. Finally, Section 4.5 summarises the main results, discusses limitations of the presented method, and lists some ideas for further work.

4.1 Introduction

4.1.1 Optical Gravitational Lensing Experiment (OGLE)

The Optical Gravitational Lensing Experiment (OGLE) is one of the longest-running variability sky surveys, regularly monitoring the densest stellar fields. It searches for microlensing phenomena in the Galactic Bulge and the Magellanic Clouds, which contain large numbers of background stars with potential to cause microlensing events during stellar transit. OGLE was first proposed by Paczynski (1991), and aims to probe the nature of dark matter. OGLE made its first observations in 1992 (OGLE-I, Udalski et al., 1992), with subsequent phases of the experiment spanning the periods of 1997-2000 (OGLE-II, Udalski et al., 1997), 2001-2009 (OGLE-III, Udalski, 2003), and 2010-present (OGLE-IV, Udalski et al., 2015), with operation suspended between March 2020 and August 2022 due to the COVID-19 pandemic. The current phase of the experiment covers over 3000 square degrees of the sky and monitors over a billion sources.

The X-ray Variables OGLE Monitoring System was introduced in 2008 (XROM, Udalski, 2008), with the intention of providing continuous, nearly real-time monitoring of the optical counterparts of a sample of around 50 X-ray sources. The work presented in this chapter is focused on the I-band photometry of Be X-ray pulsars identified in the Small Magellanic Cloud (SMC), which are included in the XROM sample. Throughout the text, I use the SXP designation introduced by Coe et al. (2005) to identify the sources. SXP (SMC X-ray Pulsar) prefix is followed by the spin period measured at the time of discovery (in the unit of seconds), for example SXP756.

Combined OGLE-II, OGLE-III and OGLE-IV I-band light curves of sources in this sample can reach almost 26 years of coverage. They show yearly observing seasons, which typically last between 230 – 280 days. Longer time gaps separate the different phases of the experiment. The cadence of observations varies between the observing seasons, starting with the median value of ∼3 d during OGLE-II, ∼2 d during OGLE-III, and ∼4 d during OGLE-IV, with occasional intervals of daily cadence. The brightness of sources can show variations of up to one magnitude, with several sources showing extreme changes over weekly, monthly, or yearly time scales (referred to as "swoopers" by Schmidtke et al., 2013).

4.1.2 Expected causes of variability in the OGLE light curves of Be X-ray binaries

As described in Section 1.2.2, we can distinguish two most prominent mechanisms causing the optical variations in BeXRBs; one relating to pulsation of the Be star, and another relating to the perturbation of the decretion disc caused by the orbiting neutron star.

Orbital perturbation is often marked by fast rise exponential decay (FRED) profile in the optical light curve, which is caused by the collision of the neutron star and the decretion disc around the Be star. Neutron stars in BeXRBs often have eccentric orbits, which can facilitate a rapid

brightening of the system upon the decretion disc disruption, and a gradual fading, as the neutron star moves away from periastron. The neutron star disruption of the decretion disc is also associated with X-ray emission, therefore the orbital period can be unambiguously determined given a periodic X-ray signal.

We also expect a more sinusoidal optical profile for systems with less eccentric orbits and/or a warped/diffuse decretion disc. In such systems, the neutron star would travel through regions with a gradually changing gas density, and the moment of collision with the decretion disc would not be defined by a sudden optical brightening.

Periodic oscillations with sinusoidal profiles, which are not associated with the orbital period of the neutron star, are thought to be caused by non-radial pulsations of the Be star. Non-radial pulsations are expected to have periods between 0.1-2 days and can therefore alias with the sampling pattern (the window function) of the light curves, creating periodogram features which can be easily confused with binary behaviour. Aliases are apparent light curve oscillations which are caused by periodic changes in the intensity of the source, which have a period that is shorter than the typical observation frequency of the light curve. Aliases are expected to appear at the frequencies given by

$$f_a = |m \times f_N - f_s|, \tag{4.1}$$

where f_N is the sampling frequency, f_s is the frequency of the signal, and m is an integer.

4.1.3 Motivations for the time-frequency analysis

The aperiodic, long-term variability of I-band photometry is mostly driven by the changing size of the Be star decretion disc. As mentioned in Section 1.2.2, the exact causes of mass loss and disc formation are not known, but it is hypothesised that non-radial pulsations could be the cause of mass ejection from the rapidly spinning Be star, resulting in disc formation. Given that the relative changes in size of the decretion disc can be directly inferred from the I-band magnitude, the hypothesis of non-radial pulsation driving mechanism can be tested via the tracking of periodic oscillations in the optical light curve. If non-radial pulsations were causing the swelling of the decretion disc, we would expect to observe that time intervals of I-band brightening are preceded by (or coincide with) periods of pulsation-driven oscillations. In order to test this hypothesis, periodic variability needs to be detected, classified as having originated from non-radial pulsations, and also be localised in time so that their occurrence can be correlated with I-band changes.

For these reason, I apply the WWZ transform, which allows for the detection of periodicities while maintaining adequate time resolution. For the purpose of signal classification, I apply the method introduced by Bird et al. (2012). They analysed the OGLE-II and OGLE-III photometry of 49 X-ray sources, and presented evidence for two types of periodic behaviour. Authors calculated two new metrics which allow for some differentiation between the different types of variability, based on the shape of the phase-folded pulse profile.

In addition to the goal of probing the physics of this specific subset of sources, the additional aim of this work is to develop a method of automated, time-resolved signal detection and classification that can test the whole range of frequencies in a single pass. As such, this method has the potential for being developed into a pipeline for the automatic processing of a large photometric data sets. New sky surveys like LSST (Ivezic et al., 2019), will be monitoring millions of sources, and the tracking of appearance and disappearance of periodic signals in the light curves of sources of interest will be essential to conduct studies of larger samples, for example the sample of BeXRBs discovered in the future.

4.2 Data preparation

In this work, I use OGLE-II data kindly provided by the authors of Bird et al. (2012), and supplemented with the data from OGLE-III and IV, which are publicly available from the OGLE website[1] as part of the XROM sub-catalogue. The OGLE-IV XROM data of sources SXP18.3, SXP264 and SXP893 are not included in the online archive, and the data presented here was kindly provided by Andrzej Udalski. The sources are not present in the XROM sample, because in this phase of the experiment, they are found in the gaps between the detectors of the 32-chip camera, outside the reference images of OGLE-IV. However, the sources serendipitously appear in the active region of the detector due to imperfect setting of the telescope. The data presented here were reduced independently of the standard reduction of the OGLE-IV photometry, using non-standard reference images that cover the whole 32-chip mosaic of the camera.

The Be star non-radial pulsations would be best observed in the V-band photometry, since the I-band is dominated by the cooler decretion disc, while the brightness changes of the star itself would be the most detectable in the V-band. However, the V-band light curves available from the OGLE survey are much sparser than the I-band data, and as such are not adequate for the periodicity search. Therefore, this analysis is limited to the I-band data only.

Incorporation of the OGLE-II data into this study requires some pre-processing to ensure that it is compatible with data of OGLE-III and IV. Comparison of OGLE-III data that was also provided by Bird et al., and the version available online[1], revealed that the two versions of the data were inconsistent. The data presented in Bird et al. (2012) were reduced independently of the more recent OGLE-XROM reduction. Different modifications of Julian Day were applied, resulting in a 0.5 day offset between the public data and data from Bird et al. (2012) for some, but not in all the sources.

In order to ensure that OGLE-II, OGLE-III and OGLE-IV data is compatible and that their time stamps are modified consistently, I calculate the timestamp-to-timestamp distance as the function of time offset between the two versions of the OGLE III light curve and find the minimum for each source, which indicates the offset for which most of the time stamps overlap. More specifically, I

[1] http://ogle.astrouw.edu.pl/

normalise the flux of the two light curves between the values of 0 and 1, graph them together, and calculate the sum of distances between consecutive data points in the time-flux plane. I offset (shift) one of the light curves in time, calculating the distances for offset values in the grid ranging from -5 to 5 days with a step of 0.1 day. I find that the light curves of some sources produce a sharp dip in the sum of point-to-point distances for offset values of 0 or 0.5 day, indicating that the data points of the two OGLE III light curves overlap. I adjust the time stamps of OGLE II data from Bird et al. (2012) using the offset values, which produced significant reduction in the point-to-point distance. I then combine this data with OGLE III and IV from the online archive[1], and use the resulting light curves in the analysis described below.

I use Prophet, an open source Python library for time series forecasting released by Facebook (Taylor and Letham, 2018), to perform light curve de-trending. Prophet can fit models with additive non-linear trends and periodic components. It is a tool that was designed to forecast business time series with interpretable parameters that allows for a semi-automated fitting without the prerequisite of the detailed knowledge of the underlying model. Prophet is a regressive model which forecasts the value of a variable y at a time t using three main components,

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{4.2}$$

where $g(t)$ is the trend function, $s(t)$ is the periodic (seasonal) component, $h(t)$ is the holiday component that is used to model irregular events, and ϵ_t is a normally distributed error term. Prophet frames the model fitting as a curve fit and not a time series fit, which allows for the fitting of an unevenly sampled time series.

Prophet allows for two modes of fitting the trend: the nonlinear saturating growth trends, and a linear trend with change points. The former method uses a logistic function to fit a saturating maximum or saturating minimum trend, and since the trends in the optical light curves of the BeXRBs tend to be more complex, I opt for fitting of a linear trend with S number of change points at times s_j, for $j = 1, \ldots, S$. The trend function if given by

$$g(t) = (k + \vec{a}(t)^{\mathrm{T}}\vec{\delta})t + (m + \vec{a}(t)^{\mathrm{T}}\vec{\gamma}) \tag{4.3}$$

where k is the base growth rate, $\vec{\delta}$ is the rate adjustment vector with S components, m is an offset parameter, $\vec{\gamma}$ is the offset adjustment vector with S components, and $\vec{a}$ is a change point indicator Boolean vector such that

$$a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j \\ 0, & \text{otherwise,} \end{cases} \tag{4.4}$$

so that the $\vec{a}(t)^{\mathrm{T}}\vec{\delta}$ term, i.e. the growth rate adjustment at time t, becomes the sum of adjustments up to that point. Components of the offset adjustment vector $\vec{\gamma}$ are set to $\gamma_j = -s_j\delta_j$ to make the trend function continuous. Prophet allows for the input of a predefined set of trend change points. I do not set any prior change point dates. Instead, I allow the algorithm to distribute the change points uniformly between the first and last observation of the light curve. I set the

number of change points S to be equal to the number of days between the first and last observation divided by 30, resulting in approximately one change point per month. The change point prior scale parameter is set to 0.1, which allows for a fairly fast gradient change from one observing season to the next, but does not allow the model to fit the fast, sharp flares, which should not be altered for the purpose of periodicity analysis.

Default seasonal components in Prophet have yearly, weekly and daily timescales to account for various patterns caused by human behaviour, and I disable the fitting of those. It is also possible to add custom seasonal components, and those can be useful when fitting persistent, periodic variability in the BeXRB light curves. Prophet fits seasonal components using a standard Fourier series

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos \frac{2\pi n t}{P} + b_n \sin \frac{2\pi n t}{P} \right) \tag{4.5}$$

where N is the number of Fourier components, a_n and b_n are the Fourier coefficients, and P is the period. For the general de-trending of light curves for the purpose of signal detection, I do not add any seasonal components to the model.

Holiday components of a Prophet model allow for the fitting of events which are predictable but non-periodic. That is accomplished through the fitting of independent regressors for each day indicated as a holiday. I disable the fitting of these for the purpose of light curve de-trending.

Model fitting in Prophet is performed through maximum a posteriori estimation using the Stan (Carpenter et al., 2017) implementation of the L-BFGS algorithm.

4.3 Periodicity detection using wavelets

4.3.1 New implementation of the weighted wavelet Z-transform

Techniques presented in the previous chapters of this thesis could be useful tools for the task of identifying and differentiating types of variability based on the light curve patterns they generate. The LSTM-VAE could be useful for the characterisation of light curve segments, and periodic variability could be detected based on the values of latent variables of those segments. However, this method requires evenly sampled data. Additionally, in the current state of development of the LSTM-VAE, it is difficult to predict whether the network would pick up on the periodicities of highest frequency unless the network was re-trained on augmented data that contains the complete range of expected signal frequencies. Finally, the LSTM-VAE would force the user to make a trade-off between the localisation of signals and the lowest frequency that could be tested for: the light curve segments create a high-pass filter, causing signals with a long period to be missed, whilst the temporal resolution of a period search would be dictated by the segment length and stride size.

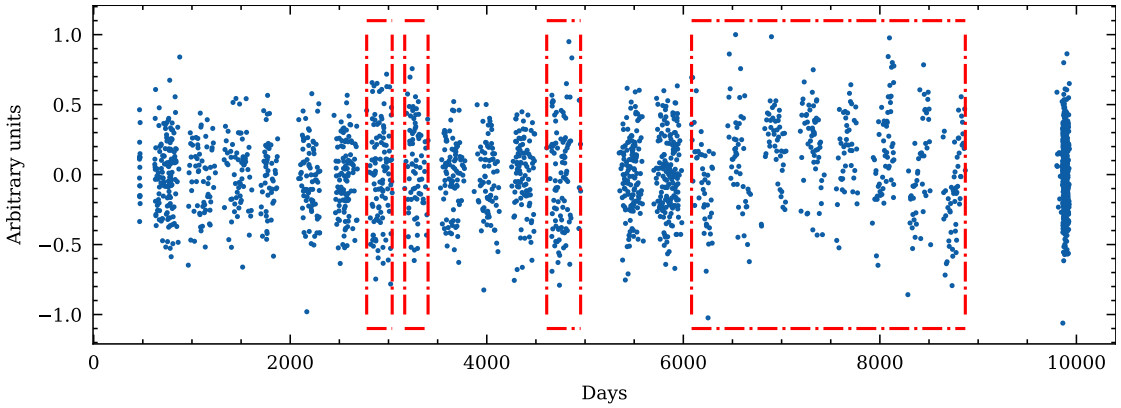
Given these considerations, I decide to use methods designed for the analysis of unevenly sampled data instead. As explained in Section 1.3.2, the use of WWZ as opposed to dynamic Lomb-Scargle spectra can be beneficial, because it does not require the setting of a window size, thanks to the non-uniform time-frequency resolution grid (illustrated by Figure 1.5). For this purpose, I develop a new Python implementation of the WWZ transform based on the code from the Python package LIBWWZ[2]. LIBWWZ is based on the Python 2.7 code by Aydin (2017), and the Fortran code by Templeton (2004), which in turn is based on the BASIC code by Foster (1996). LIBWWZ code was developed for the application in geophysical data collection systems (Garcés et al., 2022), but has been used for the analysis of several astronomic objects since then as well (Urquhart et al., 2022; Lin et al., 2023). The main reason for the development of a new Python implementation of the WWZ is the optimisation of the LIBWWZ code to minimise the computational time, and at the same time to maintain the ease of customisation of the code in the high-level Python language.

The improvements in computation speed are achieved through the use of the Numpy package for scientific computing in Python (Harris et al., 2020), and taking advantage of its vectorisation capability. However, performing vectorised operations on Numpy arrays required that they were kept in random access memory of the workstation throughout the WWZ calculation, which was prohibitive for transforms performed on light curves with thousands of observations, for millions of combinations of frequency (ω) and time shift (τ). Therefore, in addition to vectorisation, the package Cython is used to increase the speed of iterative sums performed over the observations to calculate the response of the trial functions, the S matrix, the observation weights and the weighted variation of the data (i.e. the solving of Equations (1.12) to (1.14), (1.17) and (1.23)). Cython (Behnel et al., 2011) is a Python extension compiled using C, which greatly reduces the time required to compute numerical loops. The use of Cython to perform those sums means that the subsequent operations can be performed on Numpy arrays with only two dimensions ($\omega \times \tau$) instead of three ($\omega \times \tau \times t$). The Cython portion of the implementation is also inspired by the C implementation of WWZ by Hartman and Bakos (2016). The combined effect of these optimisations results in the reduction of computation time by a factor of up to 100 when compared to the LIBWWZ implementation. Additionally, my implementation fixes a bug of the LIBWWZ code, which does not set the weighted variation of the data to zero as it loops over increasing frequency steps. The implementation is shared publicly on GitHub[3].

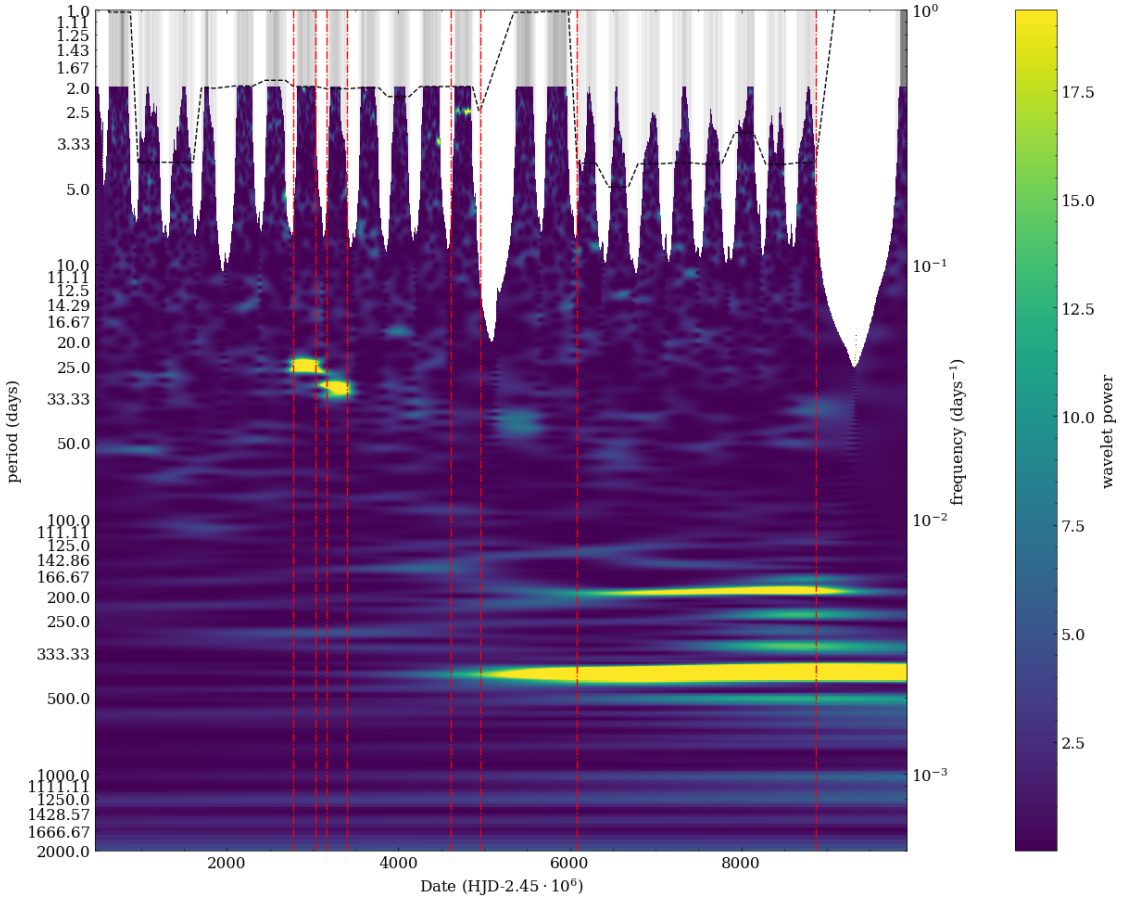
4.3.2 Parameter tuning of the wavelet transform using synthetic data

In order to test the sensitivity of WWZ and adjust the parameters of the method, I perform the WWZ-based periodicity detection on synthetic data. Figure 4.1 shows the WWZ transform of a synthetic light curve with the sampling pattern of the I-band OGLE light curve of SXP9.13, which is one of the BeXRBs from the OGLE-XROM sample. The light curve has 9462 days

[2] https://github.com/RedVoxInc/LIBWWZ
[3] https://github.com/jorwatkapola/period_search



(a) Synthetic light curve



(b) WWZ transform of the synthetic light curve

FIGURE 4.1: (a) A synthetic light curve with the typical OGLE sampling, which has transient, sinusoidal signals of 25, 30, 2.5 and 400 days injected into the intervals indicated by red dash-dotted boxes. (b) WWZ transform of the synthetic light curve. Red dash-dotted boxes indicate periodic signals, as explained above. Grey vertical lines indicate the observation timestamps. The series of horizontal dashed black lines shows the quasi-Nyquist frequency, calculated as the median time difference between time stamps in every observing season. The wavelet power value is capped at 19.5 (corresponding to $\alpha = 0.05$ for an F-distribution with ∞ and 2 degrees of freedom).

of coverage, with 2263 observations clustered into 24 observing seasons of variable length and sampling frequency (the median time separation between observations is indicated by the dashed black lines in Figure 4.1b). Signals with periods of 25, 30, 2.5 and 400 days are added to four separate sections of the light curve, indicated by red dash-dotted boxes in Figure 4.1. Signals with periods of 25 and 30 days are introduced into two consecutive observing seasons, in order to test whether WWZ can detect the relatively small change in period and localise them in time. The period of 2.5 days is added to one observing season in order to test whether a signal close to the quasi-Nyquist frequency could be unambiguously detected. A signal with a period of 400 days is introduced into eight consecutive observing seasons to test the temporal resolution of WWZ for lower frequency signals. The signals are simple sine waves, with amplitudes scaled to the standard deviation of the de-trended SXP9.13 flux. Normally distributed noise with the mean of zero and standard deviation of the SXP9.13 flux is also added to increase the scatter of the data and make the test more challenging. Figure 4.1a shows the resulting synthetic light curve.

The WWZ transform is performed on the data. In this test and analyses discussed in following sections, the numbers of time shifts and frequency steps are adjusted based on the time coverage of the light curve. One time shift is calculated for every 6.66 days of coverage, so that the number of time shifts is

$$N_\tau = \frac{3}{20\,\mathrm{days}} T \tag{4.6}$$

where T is the time coverage of the data (time difference between the first and the last observation in days). Tested frequencies range between the lower limit of 1/2000 d^{-1} and the upper limit of 1/2 d^{-1}, where the lower limit is set due to the limited time coverage (corresponding to less than 5 cycles of a 1/2000 d^{-1} signal in the longest OGLE-XROM light curve), and the upper limit is set at the quasi-Nyquist frequency of the majority of seasons in the light curve (~2 day median time gaps between the observations). The frequency step is set according to

$$\Delta f = \frac{1}{5T} \tag{4.7}$$

which corresponds to the oversampling rate of 5, as explained in Section 1.3.1. The decay rate c of the WWZ is set to

$$c = \frac{1}{20(8\pi^2)} \tag{4.8}$$

which is a factor of 20 smaller than the value of $1/(8\pi^2)$, the common choice for the light curves of variable stars, as indicated by Foster (1996). This choice is made to improve the sensitivity to high frequency signals, which would otherwise be suppressed by the increased threshold on the effective number of data points N_{eff}, which is explained in more detail below.

The WWZ has a tendency to assign extremely high power values (order of $10^5 - 10^6$) to the scalogram pixels which correspond to wavelets with frequency approaching the sampling rate (I refer to elements of the scalogram as pixels, where each pixel corresponds to a wavelet with a specific frequency f and a specific time shift τ). Extremely high power values are also

sometimes assigned to wavelets which are positioned at the edges of the observing seasons. The high power values of those wavelets make it difficult to differentiate them from real detections, so they must be filtered out. Wavelets causing false positives in WWZ tend to have small N_{eff} values, and therefore, I filter out wavelets with N_{eff}<10. This reduces the sensitivity of WWZ to high frequency signals, because high frequency wavelets do not have a long enough time span to have significant amplitude over the required number of data points. This issue is mitigated by the decreased decay constant value, which allows the wavelets to continue for several extra oscillations.

Having established the values of all of these settings, the WWZ transform of the synthetic light curve is performed. It took 4 minutes and 45 seconds to compute (for 1419 time shifts and, 23609 frequency steps). Figure 4.1 shows the resulting scalogram. Pixels with power equal to zero are masked out with white colour, and for the remaining pixels the colour bar is capped at the value of 19.5 (corresponding to the critical value of $\alpha = 0.05$ for the F-distribution with ∞ and 2 degrees of freedom; this choice is discussed in more detail in Section 4.3.3). All four signals are clearly detected, with peak frequencies of the scalogram corresponding to periods of 25.5, 29.9, 2.5 and 409.0 days. Detections corresponding to the periods of 25, 30 and 2.5 days are well constrained in time, with little power leaking into the pixels outside the corresponding time windows. They are also clearly separated in time and frequency, with a clear transition from 25 to 30-day period. The 2.5 day detections appear as two separated power peaks rather than one, and only one of them exceeds the 19.5 power level. The low power and its uneven distribution through the signal window are caused by a combination of effects; the significant amount of Gaussian noise and the signal frequency being close to the sampling frequency. However, the detection can be clearly differentiated from any spurious peaks due to its extent, which indicates that wavelets of this frequency produced high power over multiple time shifts. The signal with the longest period is not constrained in time as well as the other signals; a significant amount of power leaks outside the time window where the signal is present. The reduced temporal resolution of the low frequency signal is a consequence of the reduced decay constant value. However, the power distribution over time shows a Gaussian-like profile, with the peak at the centre of the time window where the 400 d signal was injected. A signal with a period of 191.0 days is also detected in the same time window as the 400-day signal, and it is caused by the window function of the light curve; power is also found at the 191 d period if all the flux values are set to 1.

Figure A.6 shows a WWZ transform of the same data, using alternative settings. There, the decay constant is set to the conventional value of $c = 1/(8\pi^2)$, which is 20 times larger than the one used for the transform shown in Figure 4.1. No filtering based on the value of N_{eff} is applied. Comparison of the two figures illustrates that the smaller decay constant can provide a more complete coverage of the high frequency end of the scalogram, and together with the N_{eff} filter, it removes many of the spurious detections of individual pixels showing extremely high power. These settings are therefore required to detect the signal with 2.5-day period, which is otherwise much more difficult to distinguish from the spurious peaks in the scalogram.

4.3.3 Detection and characterisation of signals

Following the WWZ transform, the pixels corresponding to likely detections must be identified in order to undergo further characterisation. However, estimation of the WWZ power significance level is not straightforward. Despite significant improvement in the speed of the WWZ code, evaluating thousands of Monte Carlo simulations for every light curve would be prohibitively expensive in terms of computation time. Instead, I identify the detection power threshold based on the F-distribution critical value. As Foster (1996) remarked, The WWZ power should follow the F-distribution with ($N_{\rm eff}$−3) and 2 degrees of freedom, but since $N_{\rm eff}$ is not an integer, the actual distribution of the WWZ power cannot be precisely defined.

However, the effect of increasing the first parameter of the F-distribution (i.e. the numerator degrees of freedom in Equation (1.21), df_1) does not have a significant effect on the critical value. For example, the smallest value of ($N_{\rm eff}$−3) observed in a WWZ with $N_{\rm eff} < 10$ pixels filtered out is 7, and the $\alpha = 0.05$ critical value of an F-distribution with 7 and 2 degrees of freedom is 19.3532. As the first parameter is increased, the critical value asymptotically approaches the maximum of 19.4957 at $df_1 = \infty$. Therefore, for the practical reason of identifying detections in the scalogram, it is reasonable to use the F-distribution with ∞ and 2 degrees of freedom, which provides the highest critical value for any value of α. An inspection of the WWZ transform of the synthetic light curve shows that the real signals create power peaks reaching power values between 31-124. However, multiple repetitions of this test with re-sampled Gaussian noise reveal that the power of the 2.5 d peak can drop to ∼20. Hence, the detection threshold is set at 19.5, corresponding to $\alpha = 0.05$, which allows for all the true signals in the test light curve to be detected.

This WWZ power threshold is then used to identify the pixels associated with candidate detections. In the time shift-frequency plane of the WWZ scalogram, the detections appear as regions of high power with one or more peaks, and the power decays gradually, in both frequency and time shift directions, as the distance from the peak increases. This results in multiple neighbouring pixels exceeding the detection power threshold. For the purpose of signal characterisation, pixels which exceed the detection threshold due to the same signal frequency and at a similar time, must be clustered and analysed together. I process each cluster separately, in order to be able to differentiate signals of different physical origin appearing at different times. I identify the pixels associated with each disconnected cluster using the agglomerative clustering algorithm with the single-linkage method (see Section 1.4.4.2 for more details about the algorithm), and apply it to the Boolean detection matrix derived from the WWZ scalogram. The distance threshold of the clustering algorithm is set to 1.1, because neighbouring pixels are separated by the distance of 1, whereas the shortest possible distance between non-neighbouring clusters is 2.

For every cluster of pixels above the detection threshold, I calculate the local WWZ spectrum by summing the power over the pixels spanned by the cluster, in the time shift direction. In order to account for the possibility of multiple signals being present in the cluster, I use a peak finding algorithm included in the SciPy package (Virtanen et al., 2020) to reveals the most prominent

frequencies in the cluster. The algorithm finds the local maxima in a series of values through simple comparison of neighbouring values. As an additional requirement, I set the prominence parameter to 10% of the maximum power value in the spectrum. The prominence of a peak is defined as the minimum drop in elevation between the peak and a power value higher than that peak (i.e. the vertical distance between a peak and its key col). Setting a prominence requirement prevents the algorithm from marking all the local maxima as peaks, so it only returns ones which stand out significantly.

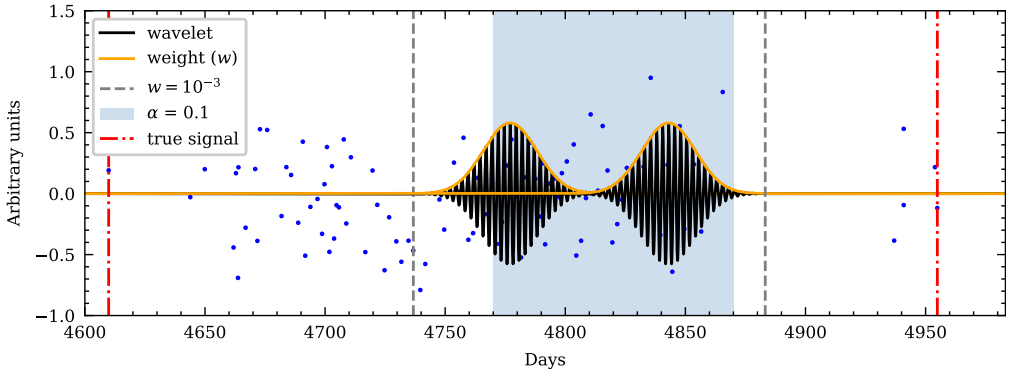
For every frequency peak of a cluster, I compute wavelets of the corresponding frequency at the first and last time shift of the cluster (using the wavelet function from Equation (1.20)). Then, using these two most extreme wavelets, I find the most extreme time values with the wavelet weight above the minimum threshold of 10^{-3} (using the weight formula from Equation (1.19)). These timestamps define a time window where observations are deemed to have a significant contribution to the power of the cluster. This approach is used to maximise the recovery of observations which contribute to the power within the cluster, which makes the phase-folding analysis better defined, particularly for high frequency signals, which can be detected using a relatively small number of observations. However, in case of the low frequency clusters, it also results in a window which stretches far outside the cluster. I therefore add another constraint on the window of observations, restricting it to the first and last τ where the pixel power is larger than the critical value of $\alpha = 0.1$. Additionally, if no observations are found between the first and last τ exceeding the critical value of $\alpha = 0.05$, I deem that to be a false detection. Figure 4.2 shows an example illustration of all of these detection criteria, overlaid on top of the sections of the light curve where 2.5 d, 25 d and 400 d signals are detected. The final sets of observations which are used in the phase folding analysis are marked by the blue shaded regions.

For every cluster above the detection threshold, and every frequency peak within it, I phase-fold the observations matching the detection criteria described above, and then bin the phase-folded data. The number of bins is a set depending on the number of observations, according to

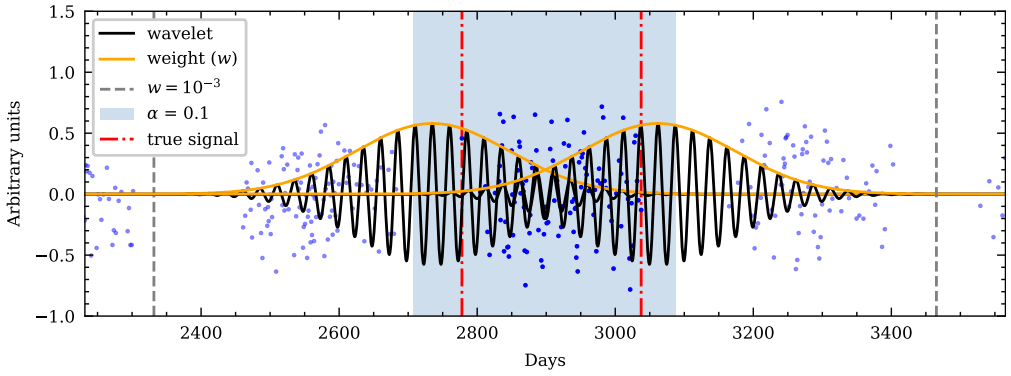
$$\#\text{bins} = \frac{30 \times \#\text{observations}}{1000}, \tag{4.9}$$

with the minimum capped at 10 and the maximum capped at 30. The number of bins is automatically adjusted for each subset of observations undergoing the phase-folding analysis. This is because an increase in the number of bins can improve the resolution of the phase-folded profile, especially when there are many observations available. On the other hand, reducing the number of bins can prevent the fitting of sparse profiles with many empty bins, which is useful for cases where there are only a few observations available.

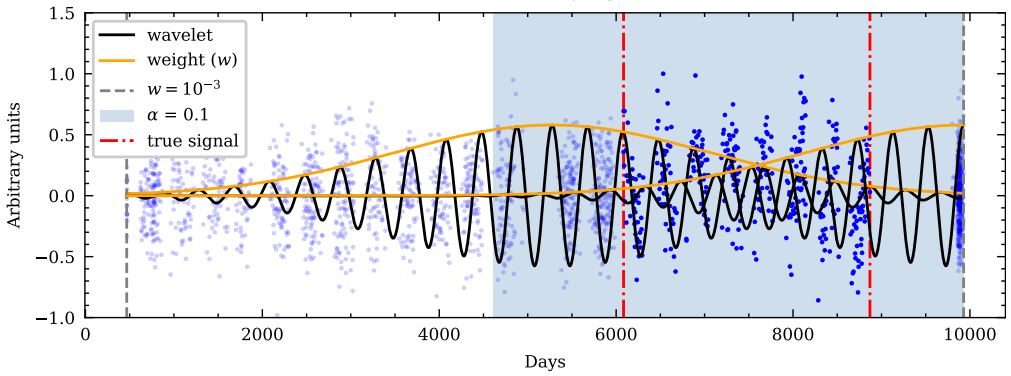
Phase span and phase asymmetry metrics are calculated for every phase-folded profile. The metrics were introduced by Bird et al. (2012) for the purpose of characterisation of phase-folded pulse profiles, with the aim of differentiating sinusoidal profiles from FRED-like profiles. The phase span (PS) is defined as the full width at half maximum of the peak in phase space, while the phase asymmetry (PA) is defined as the ratio of phase differences between the maximum and



(a) detection of 2.5 day signal



(b) detection of 25 day signal



(c) detection of 400 day signal

FIGURE 4.2: Example sets of observations identified as signal detections in the WWZ of the synthetic light curve, shown in Figure 4.1, together with illustrations of the detection criteria. Bounds of the window where the signal was injected into the light curve are marked with red dash-dotted lines, like in Figure 4.1, and observations which had the signal added to them are marked by solid blue points. The final detection window is marked with the blue shaded region (bounded by the first and last pixels with power corresponding to at least $\alpha = 0.1$, and within the window where the wavelet weight of the drawn wavelets is at least $w = 10^{-3}$). The wavelets are drawn at the first and last τ of the cluster with power corresponding to at least $\alpha = 0.05$. The sub-figures show three detection clusters and the wavelets corresponding to signals with periods of (a) 2.5 d, (b) 25 d, and (c) 400 d.

the points where the profile drops to 10% of the maximum. In addition to this, I also find an alternative PA metric, by calculating the ratio of phase differences between the edges of the peak at the half-maximum and at the 10% of the maximum.

However, calculating PS and PA metrics directly from the binned data can be unreliable, because the profiles do not always decay monotonically, either due to noise or sparseness of data. In order to approximate these metrics in a more reliable manner, I first smooth the binned profiles by fitting the Savitzky-Golay filter to the series of mean values of the binned data. Savitzky-Golay filter is applied by locally fitting a least-squares polynomial to an evenly sampled series, using a moving window (see Schafer, 2011, for more details about the filter). I use the SciPy (Virtanen et al., 2020) implementation of the filter with the polynomial order of 2. The window size of the filter is adjusted to the number of bins, according to

$$\text{Savitzky-Golay window length} = \frac{\#\text{bins}}{2}, \tag{4.10}$$

so the window length can vary between 5 and 15, depending on the number of phase bins.

Figure 4.3 shows the results of the phase-folding analysis of the synthetic light curve shown in Figure 4.1. The sub-figures show the binned, phase folded data, with heights of bars representing the mean flux in each bin. The black lines show the Savitzky-Golay filter fit to the mean values of binned data. The darker shaded region corresponds to the full width at half maximum, and the lighter region shows where the profile is above 10% of the maximum. The PA and PS metrics calculated for the detection of true signals place all of them in the sinusoidal profile region of the PS-PA plane (with approximate boundaries at PS>0.4 and PA<2 according to Bird et al. (2012)).

The profile of the 29.9 d period (Figure 4.3c) has the PA of 2.135, which is slightly higher than the value expected for a sinusoid profile. However, upon inspection it becomes clear that the 29.9 d pulse does have the wide, symmetric profile of the sine wave, and that the PA value is inflated due to the slightly skewed position of the maximum. The alternative PA (calculated at the half-maximum level) has the value of 1.423, and is a more accurate representation of the symmetry of the profile.

Figure 4.3f shows the profile of a spurious detection found at around 4770 d timestamp, outside any window where signals were injected. Together with a second spurious detection found at around 8480 day time stamp, they are both caused by the Gaussian noise. Both of these spurious detections produce sparse phase-folded profiles due to the small number of contributing observations, resulting in a significant number of empty bins. I filter out many spurious detections like this by setting a requirement that at least half of the bins must be populated.

Results of this test with synthetic data show that it is possible to automatically detect signals using the WWZ transform, and to characterise the peaks using the PS and PA metrics, but care must be taken in case of high frequency, poorly-sampled signals.

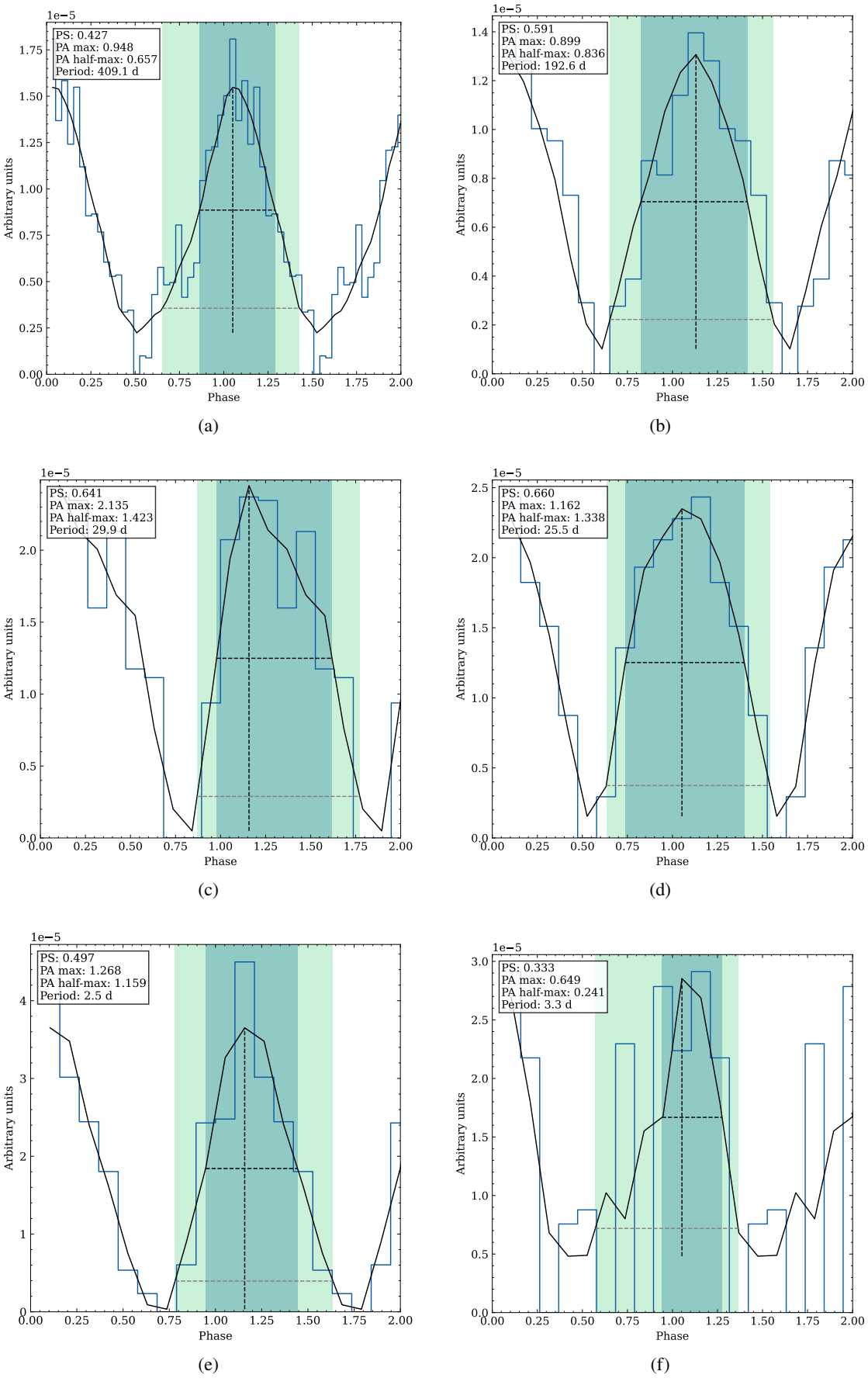


FIGURE 4.3: Phase folded profiles of data responsible for detections presented in Figure 4.1. The black lines show the Savitzky-Golay filter fitted to the mean values of binned data. The vertical black dashed line indicates the maximum, and the horizontal black dashed line shows the half-maximum level. The grey dashed line indicates the 10% of the maximum level.

TABLE 4.1: Sources included in the WWZ analysis. All listed sources come from the OGLE-XROM sample of Small Magellanic Cloud BeXRBs. Dates are presented as the Truncated Julian Date (TJD), which is defined as the Heliocentric Julian Date (HJD) minus 2.45×10^6 days. Candidates for newly detected periodicities are listed together with the classification of the profiles derived from the phase-folded light curves.

Source name	No. observations	TJD range	New candidate periods (days)
SXP0.09	2542	627.9 – 9928.5	
SXP0.92	2485	625.9 – 9928.5	
SXP2.763	1425	2086.9 – 9916.6	
SXP9.13	2263	466.5 – 9928.5	Sinusoidal (3.0 or 0.753)
SXP22.07	1048	2104.9 – 9910.6	FRED-like (146.4)
SXP25.5	2254	621.8 – 9928.5	
SXP31.0	1252	2104.9 – 9918.5	
SXP34.08	1822	2086.9 – 9928.5	
SXP46.6	2913	2086.9 – 9928.5	
SXP59.0	2840	626.9 – 9928.5	
SXP74.7	2509	466.5 – 9928.5	
SXP91.1	1734	2086.9 – 9928.5	
SXP101	1578	2086.9 – 9927.5	Sinusoidal (3.0 or 0.750)
SXP138	2045	2086.9 – 9928.5	
SXP152.1	1968	2086.9 – 9928.5	
SXP169.3	1995	2086.9 – 9928.5	
SXP264	1199	621.8 – 7258.8	
SXP293	1249	621.9 – 7367.6	
SXP304	2215	627.9 – 9928.5	FRED-like (341.3)
SXP323	2246	621.8 – 9928.5	
SXP327	2046	2086.9 – 9928.5	
SXP504	2293	621.8 – 9928.5	
SXP565	2127	628.9 – 9928.5	Sinusoidal (4.9 or 0.831)
SXP701	2311	621.8 – 9928.5	
SXP726	2098	627.9 – 9928.5	FRED-like (112.0)
SXP756	2186	466.6 – 9928.5	
SXP893	1195	466.5 – 7364.6	
SXP1323	2178	627.9 – 9928.5	

4.4 Wavelet-based detection of signals in the OGLE-XROM photometry

In this section, I discuss the results of applying the WWZ-based periodicity detection pipeline to the I-band light curves of a sample of OGLE-XROM sources listed in Table 4.1. The light curves of the sources are de-trended using the trend component of the PROPHET model. Once the trend is removed, the WWZ transform is performed on the data. As described in the previous section, the range of tested frequencies is bound by the lower limit of 1/2000 d^{-1} and the upper limit of 1/2 d^{-1}, with the frequency step adjusted according to Equation (4.7). I limit the discussion to the most prominent detections from the WWZ transform of each source, however the scalograms

with all detections can be found in Appendix A, together with light curves of the analysed sources, which are not included in this section.

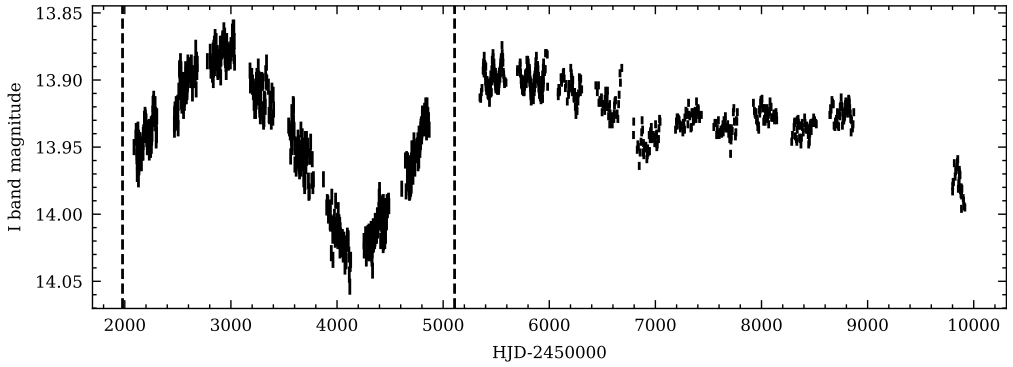
4.4.0.1 SXP0.09

The WWZ of the SXP0.09 light curve shows two separate detections of the ~17 d period, first with the period of 17.0 d (20.6 power), centred at 3496.8 TJD, and second with the period of 17.3 d (42.3 power), centred at 9791.8 TJD, both lasting for about one observing season. Some power at the same period can also be observed around 6000 TJD, however the power peak is below the detection threshold. Other short detections are also found at 4891.2 TJD with the period of 168.3 d (22.3 power), and at 3496.8 d with the period of ~13 d. The latter has a multi-peaked spectrum, with the greatest power at the period of 13.9 d (20.0 power). Schmidtke et al. (2013) report a detection of a 13.3 d period in the same observing season, and they interpret it as an alias of a 0.93 d pulsation. Power at period ~250 d (below detection threshold) is also present throughout the OGLE-III seasons, as reported by Rajoelimanana et al. (2011) and Schmidtke et al. (2013), but it ceases at around 6000 TJD. PS and PA values of all signals indicate sinusoidal profiles.

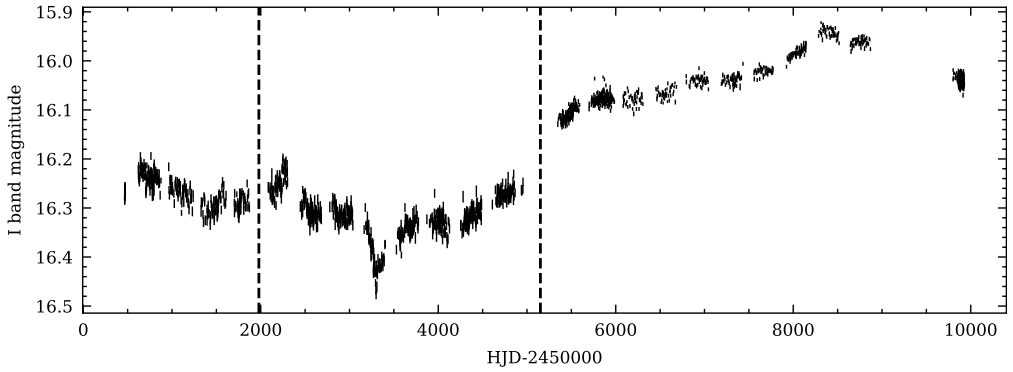
4.4.0.2 SXP2.763

A signal with a period of ~82 d is present across almost the entire light curve of SXP2.763, with three detection clusters stretching between 2086-3448, 4569-6612 and 8080-8835 TJD (showing peak powers of 63.9, 406.4 and 26.0 respectively). There is also a fourth detection cluster between 7440-7667 TJD (20.6 power), where the signal seems to temporarily drift to a period of 86.9 d. This apparent drift is observed during a period when the amplitude of the oscillations observed in the light curve decreased (see Figure 4.4a). The amplitude of oscillations is correlated with the brightness of the source; they are the most prominent in the brightest parts of the light curve, and disappear completely in the dimmest parts. Hence, the power of the ~82 d detections also correlates with the brightness of the source. Phase-folded profiles of the ~82 d detections all have a PA$\leq$1; the two most prominent detections however have PS of 0.38 and 0.34, which places them outside the sinusoidal region of the PS-PA plane. The narrow profile of the oscillations and the amplitude-brightness correlation point towards the orbital origin of the signal. Orbital origin is also suggested by Schmidtke et al. (2006) and Coe and Kirk (2015). The period of 81.81 ± 0.06 was reported by Bird et al. (2012).

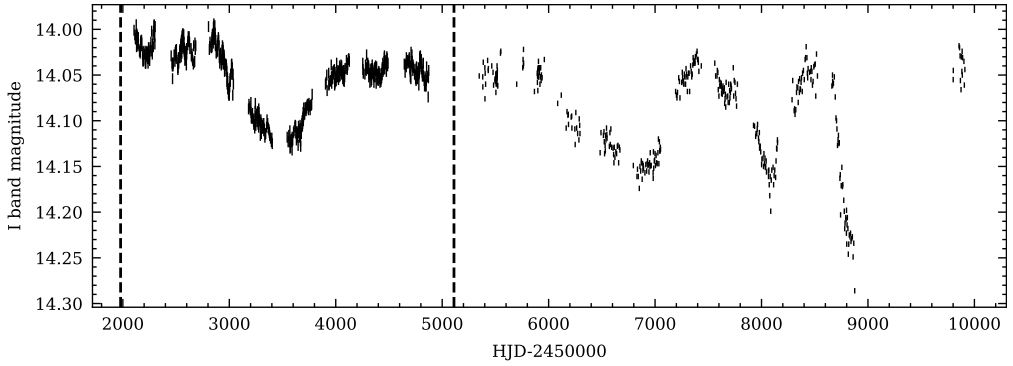
Four other detections are found in the WWZ of SXP2.763. This includes a peak at 3502 TJD at a period of 42.1 d with power of 24.6, which is a potential harmonic of the ~82 d signal. Three peaks corresponding to periods of 131.1 d, 148.0 d and 116.7 d are found in the last observing season and are likely spurious, resulting from fits to the 29 isolated observations.



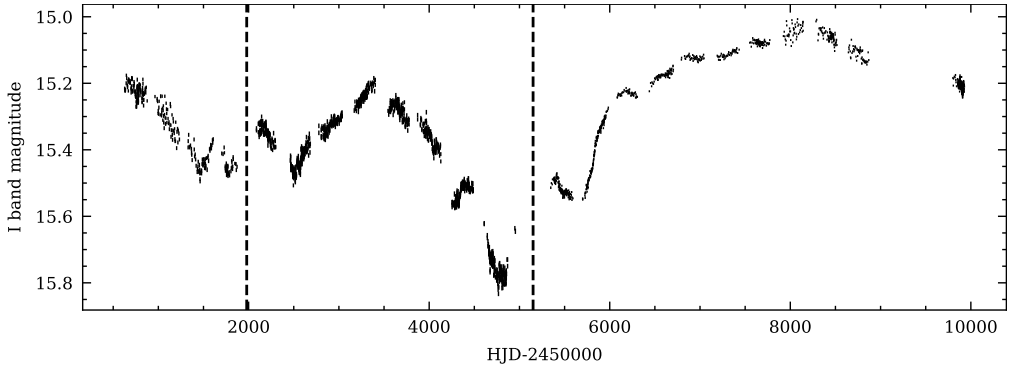
(a) SXP2.763



(b) SXP9.13



(c) SXP22.07



(d) SXP59.0

FIGURE 4.4: I-band OGLE light curve of analysed sources. OGLE-II, III and IV are separated by black dashed lines.

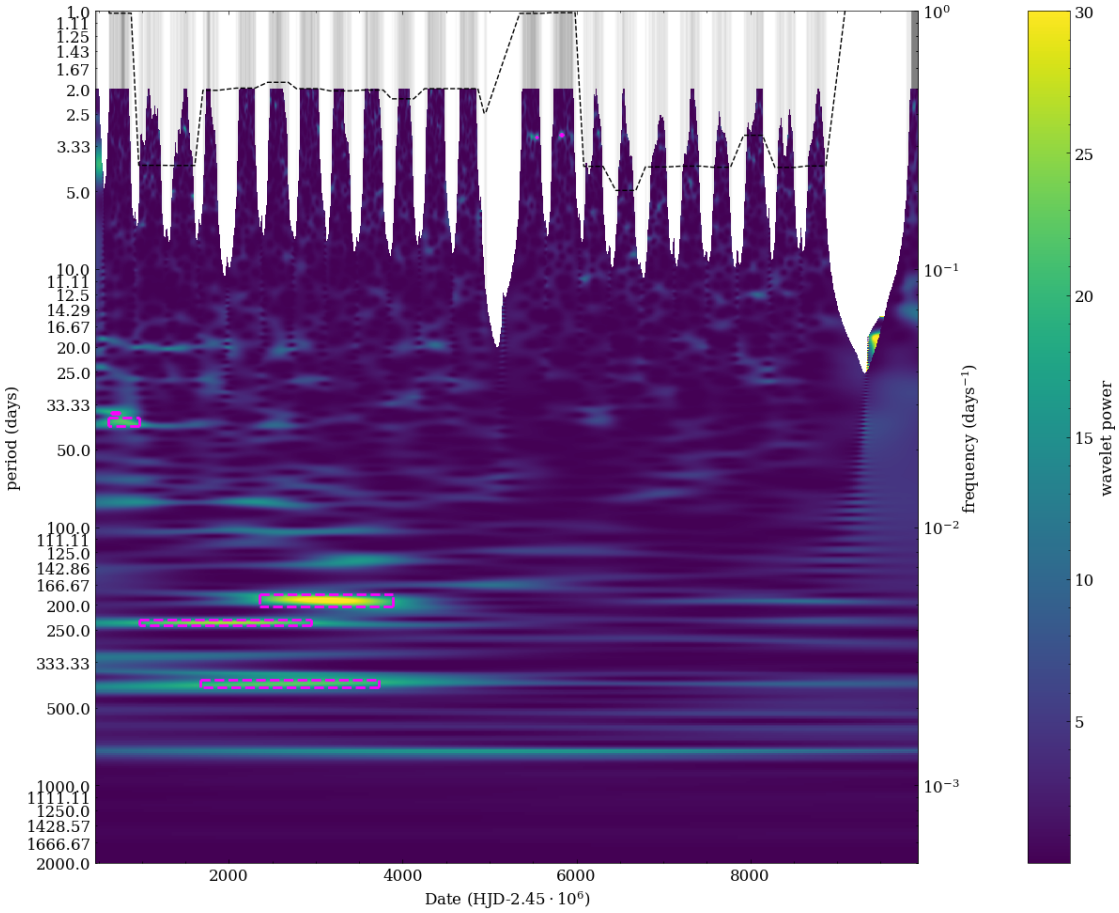


FIGURE 4.5: WWZ transform of SXP9.13 light curve. Grey vertical lines indicate the observation timestamps. The series of horizontal dashed black lines shows the quasi-Nyquist frequency, calculated as the median time difference between time stamps in every observing season. The wavelet power value is capped at 19.5 (corresponding to $\alpha = 0.05$ for an F-distribution with ∞ and 2 degrees of freedom). Dashed magenta boxes indicate the most extreme pixels which exceed the $\alpha = 0.05$ power level, for every detection.

4.4.0.3 SXP9.13

Detections in the WWZ of this source in the order of decreasing period are: 405.6 d (22.3 power, 1774.4-3576.1 TJD), 234.6 d (26.0 power, 973.7-2935.5 TJD), 192.6 d (32.8 power, 2361.6-3876.3 TJD), 39.4 d (22.1 power, 620.0-927.0 TJD), 36.0 d (19.8 power, 653.4-733.5 TJD). A 3.0 d period is also detected in the first two observing seasons of OGLE-IV (23.3 power, 5818.1-5831.5 TJD, and 20.9 power, 5537.9-5544.5 TJD), shown in Figure 4.5. The most prominent detection has the period of 192.6 d, and it is caused by the window function of the light curve; it was also detected in the synthetic light curve test described in Section 4.3.2, although it peaked at a different time. Other detections with periods >100 d have noisy phase-folded profiles, and have PS-PA values that indicate sinusoidal profiles.

Signals with ~40 d period were previously reported by Edge (2005) and Galache et al. (2008), who also reported power at ~20 d. The ~20 d period was in turn interpreted as an alias of a 0.95 d non-radial pulsation. However, Rajoelimanana et al. (2011) interpreted it as a harmonic

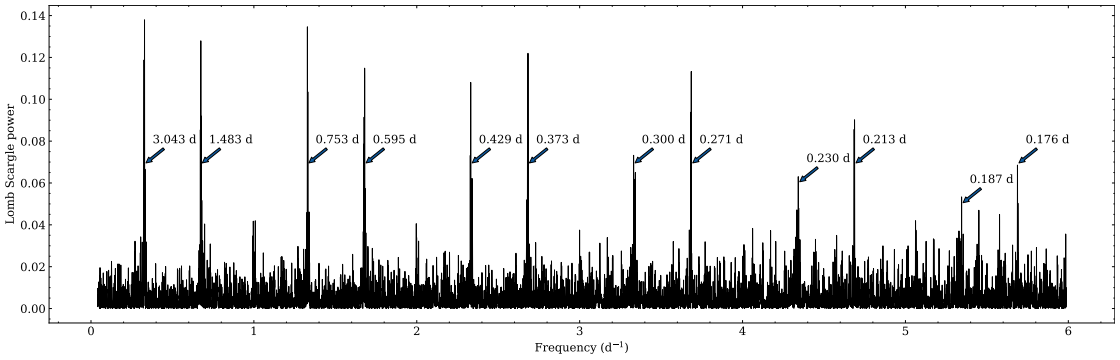


FIGURE 4.6: Lomb-Scargle periodogram of a section of SXP9.13 light curve between 5346-5985 TJD, over the range of frequencies expected for a non-radial pulsation of the Be star. The most prominent peaks are labelled with their corresponding periods.

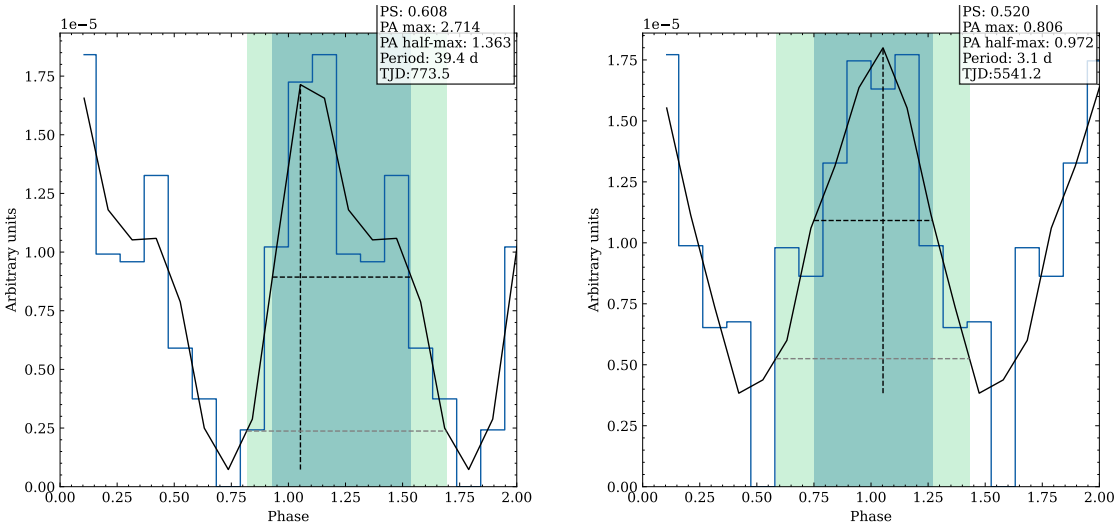
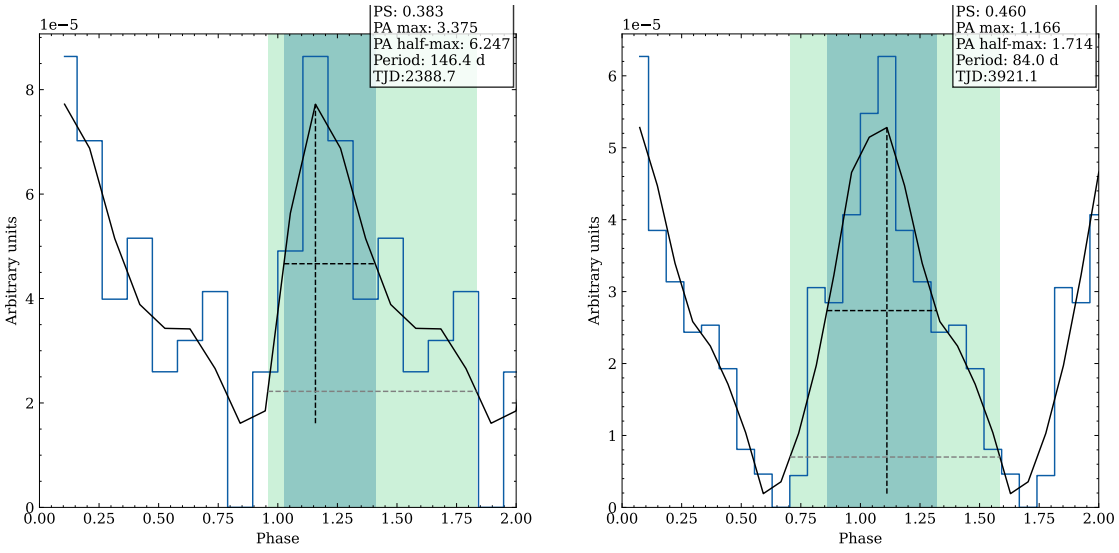


FIGURE 4.7: Phase-folded profiles of signals detected in the light curve of SXP9.13.

of a 80.1 ± 0.06 d period, together with other harmonics at 20 and 26.6 d. Furthermore, Galache et al. (2008) reported an X-ray orbital period of 77.2 ± 0.3 d. However, the ~80 d period was not found by Schmidtke et al. (2013). I find some power below the detection threshold at ~20 d and ~80 d periods, which together with the asymmetric profile of the 39.4 d period (PA=2.7, see Figure 4.7a), point toward the binary orbital period as the cause of the signal. Furthermore, Coe and Kirk (2015) gave the X-ray period of 40.10 d. However, these signals seem to vanish after 3000 TJD, which coincides with a dip in the brightness of the source by ~0.2 mag (see Figure 4.4b), and they do not re-appear following the re-brightening of the source, which comes back to the brightness level from before the dip, and then continues to gradually brighten up by another ~0.35 mag. These observations do not seem to support the interpretation that the ~80 d period is driven by the orbital period, unless the decretion disc becomes so diffuse following the source's brightening that its interaction with the neuron star stops being observable in the photometry, but this diffusion would also need to happen without any transition period when the ~80 d period could reappear.



(a) (b)

FIGURE 4.8: Phase-folded profiles of signals detected in the light curve of SXP22.07.

The 3.0 d period is detected in the first two seasons of OGLE-IV, with some power also appearing in the third season (6250 TJD), as well as the third season of OGLE-II (1750 TJD). I have not found previous reports of this periodicity. The phase-folded profile of the second detection is sparse, with multiple gaps, however the profile of the first detection has a clearly sinusoidal shape (see Figure 4.7b). Figure 4.6 shows a Lomb-Scargle periodogram of the first two observing seasons of OGLE-IV in the range of frequencies between $0.167 - 25\mathrm{d}^{-1}$. The most prominent peaks are found at the periods of 1.483, 0.753, 0.595, 0.429, 0.373 and 0.271 d, and they correspond closely to the alias pattern expected for an NRP with the period of 0.753 d; the aliases of such a signal are expected to have periods of 3.043, 1.489, 0.598, 0.374, 0.272, 0.214, and 0.176 d, for the nominal sampling frequency of 1 day (derived from Equation (4.1) with $m = \{1 \ldots 7\}$). Therefore, the 3.0 d period is most likely an alias of an NRP with the period of 0.753 d. The periodogram also shows peaks at the periods of 0.429, 0.300, 0.230 and 0.187, which can be aliases of the 1.483 d period (expected at 3.07, 0.754, 0.43, 0.301, 0.231, and 0.188 d).

4.4.0.4 SXP22.07

The WWZ scalogram of SXP22.07 produced detections at 271.9 d (25.1 power, 2104.9-2959.6 TJD), 238.7 d (28.5 power, 2104.9-3907.7 TJD), 146.4 d (29.2 power, 2104.9-2672.5 TJD), 108.3 d (32.6 power, 4295.0-5456.9 TJD), 84.0 d (27.1 power, 3106.5-4735.7 TJD), 65.4 d (22.5 power, 2438.8-2699.2 TJD) and 57.7 d (21.9 power, 2639.1-2866.1 TJD).

The phase-folded profiles of the pulses with 271.9 and 238.7 d are noisy, with the mean intensity varying widely from one bin to the next, which is likely an indication that the detections are spurious, potentially caused by the sampling pattern of the light curve. The power of the 238.7 d

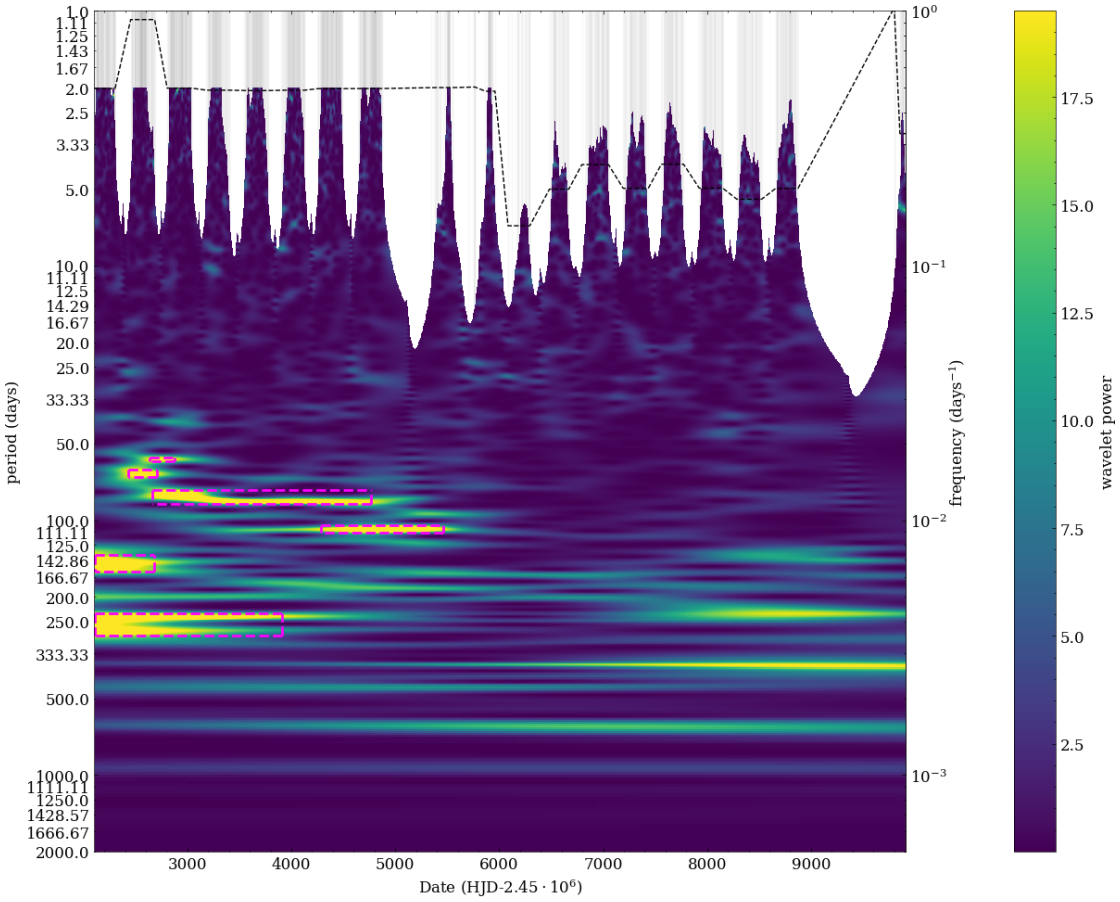
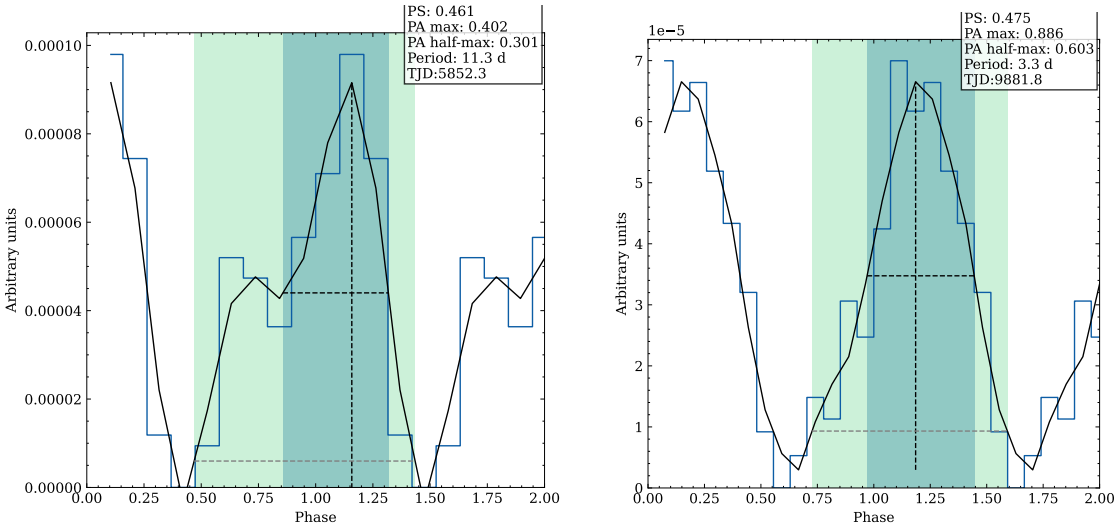


FIGURE 4.9: WWZ transform of SXP22.07 light curve. Grey vertical lines indicate the observation timestamps. The series of horizontal dashed black lines shows the quasi-Nyquist frequency, calculated as the median time difference between time stamps in every observing season. The wavelet power value is capped at 19.5 (corresponding to $\alpha = 0.05$ for an F-distribution with ∞ and 2 degrees of freedom). Dashed magenta boxes indicate the most extreme pixels which exceed the $\alpha = 0.05$ power level, for every detection.

period drops close to the background level around ~5000 TJD, and it increases again after ~7000 TJD.

Rajoelimanana et al. (2011) suggested a period of 75.97±0.06 d, and Bird et al. (2012) reported a signal with 83.7±0.1 d period, while Schmidtke et al. (2013) interpreted their results as either a ~77.3 d period or a ~0.986 d non-radial pulsation, with the latter increasing from 0.983 to 0.989 d from season to season. However, I find that a drift to a longer period can also be observed for the ~80 d signal in the WWZ scalogram, during the OGLE-III phase (see Figure 4.9). The signal drifts from 80.1 to 84.7 d period, and is also preceded by detections of 65.4 and 57.7 d periods, which mostly cover the range of periods expected from aliases of NRPs reported by Schmidtke et al. (2013). Signals with 57.7, 65.4 and 84.7 d periods have sinusoidal profiles (see Figure 4.8b for an example), and therefore it is possible that they are aliases of the NRP reported by Schmidtke et al. (2013).

The 146.4 d period has the FRED-like profile (see Figure 4.8), with PS 0.383, PA 3.375, but it is only detectable until ~3250 TJD, which coincides with the first dip in the brightness of the



(a) (b)

FIGURE 4.10: Phase-folded profiles of signals detected in the light curve of SXP25.5.

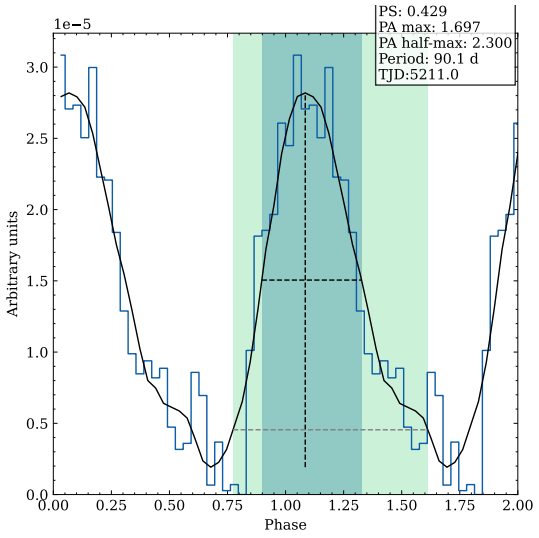
source. The brightness of the source partially recovers after the dip, but does not reach the same levels as in the first three observing seasons of OGLE-III (see Figure 4.4c). The stopping of the FRED-like oscillation when the source fades could be an indication of the interaction of the neutron star with the disc.

4.4.0.5 SXP25.5

The WWZ scalogram of SXP25.5 produces detections at 610.1 d (23.5 power, 621.8-1816.0 TJD), 276.5 d (21.6 power, 621.8-1635.9 TJD), 211.3 d (25.7 power, 621.8-1976.1 TJD), 119.2 d (21.8 power, 2269.7-2796.7 TJD), 11.3 d (23.7 power, 5778.9-5925.6 TJD), 3.3 d (218.7 power, 9835.1-9928.5 TJD), 3.0 d (22.7 power, 2202.9-2209.6 TJD).

Rajoelimanana et al. (2011) reported the orbital period of 22.50±0.01 d and Bird et al. (2012) reported 22.53±0.01 d, while Schmidtke et al. (2013) did not find this period in their analysis. Coe and Kirk (2015) gave the X-ray orbital period of 22.5 d. The WWZ scalogram shows some power at this period during the latter half of the OGLE-III phase, as well as the second season of OGLE-IV, but always below the detection threshold. The detection of an 11.3 d signal however could be a harmonic of the ~22 d signal, with some power at 7.5 and 5.6 d periods as well. The folded profile is highly asymmetric, with PA of 0.40 (left-skewed profile, see Figure 4.10a).

A signal with a period of 3.3 d is strongly detected in the last observing season of OGLE-IV, and also with the period of 3.0 d in the first season of OGLE-III, but power at the ~3 d period is also observed in other seasons of OGLE-II and III. Schmidtke et al. (2013) reported the detection of a 0.598 d NRP in OGLE-II and the first four seasons of OGLE-III. I find power at the ~3 d period in five of those seasons, and it also disappears completely in the first two seasons of OGLE-IV, just like the NRP described by Schmidtke et al. (2013). The phase-folded profile of



(a)

FIGURE 4.11: Phase-folded profiles of signals detected in the light curve of SXP31.0.

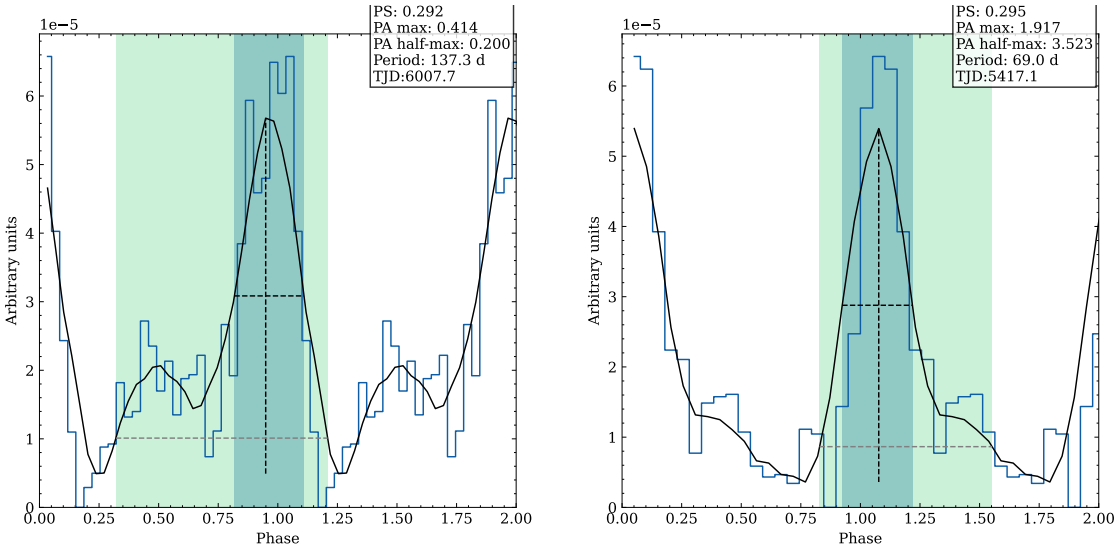
the detection at 2206 TJD has a very sparse profile, with many empty bins, but the profile of the 3.3 d detection at 9881 d has a well-defined, sinusoidal profile (PS=0.475, PA=0.886, see Figure 4.10b). Therefore, the ~3 d period is most likely an alias of the 0.598 d NRP.

4.4.0.6 SXP31.0

The light curve of SXP31.0 shows prominent outbursts throughout OGLE-III and IV, detected in the WWZ with the period of 90.1 d (80.9 power, 2104.9-8317.1 TJD). Schmidtke et al. (2013) reported a period of 90.57±0.09 d and described it as having a FRED-like profile. I find that the phase folded profile has a PS of 0.420 and PA of 1.697, which would place it in the sinusoidal region of the PS-PA diagram. However, the profile is clearly asymmetric (see Figure 4.11a), and with the alternative PA of 2.3, it is a clear indication of the FRED-like profile. Furthermore, Coe and Kirk (2015) gave the X-ray orbital period of 90.4 d.

4.4.0.7 SXP46.6

The WWZ scalogram of SXP46.6 shows two periods which persist through multiple observing seasons; the first one at 137.3 d (61.4 power, 2086.9-9928.5 TJD), and the second one at 220.8 d (54.4 power, 2180.3-6565.0 TJD). The former is detected throughout the light curve, and it corresponds to the X-ray period reported by Galache et al. (2008) at 137.36 ± 0.35 d and Kennea et al. (2018b) at 143.29 ± 4.5 d, as well as the optical periods reported by Schmidtke et al. (2007) (138.4 ± 0.9 d), McGowan et al. (2008) (~137 d), Rajoelimanana et al. (2011) (136.4 ± 0.2 d), and Bird et al. (2012) (137.4 ± 0.2 d). McGowan et al. (2008) found that the phase-folded profile of the modulation is highly non-sinusoidal, and has a double-peaked structure. I find a similar result from the WWZ detection at 137.3 d; the profile of the peak is narrow and highly



(a) (b)

FIGURE 4.12: Phase-folded profiles of signals detected in the light curve of SXP46.6.

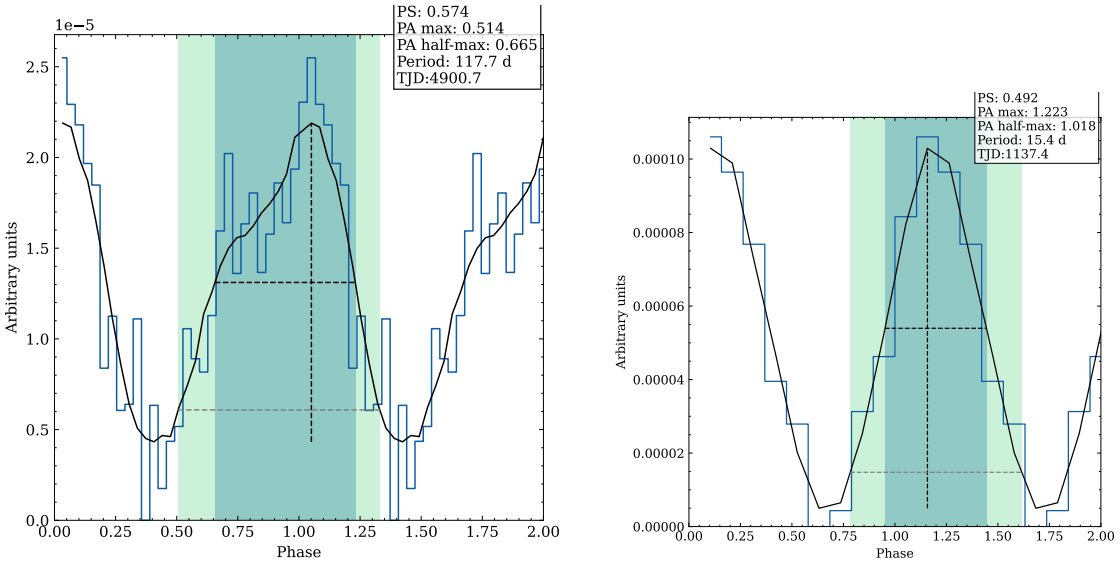
asymmetric (PS= 0.292, PA= 0.414), as illustrated in Figure 4.12a. The 220.8 d period lasts from the start of OGLE-III until ~6500 TJD and has a noisy, wide, asymmetric profile (PS= 0.483, PA= 0.640). It was also found by McGowan et al. (2008) in the Lomb-Scargle periodogram, and it can be interpreted as the beat period between the 137.3 d period and the annual observing cycle, while Coe and Kirk (2015) gave the X-ray orbital period of 137.4 d.

Another signal with a FRED-like profile (see Figure 4.12b) is detected starting from the end of OGLE-III, with period of 69.0 d (30.6 power, 4769.7-6064.4 TJD), with some power present throughout OGLE-IV. I find some power at this period during the first half of OGLE-III as well. A period of 69.2 d was also reported by Schmidtke et al. (2007) and at 68.98 d by Schmidtke et al. (2013). It is most likely a harmonic of the 137.3 d period.

The WWZ scalogram contains one large detection cluster in the last observing season of OGLE-IV, which spans the periods between ~17 d and ~160 d, with main peaks at 47.4 d (285.7 power) and 26.7 d (150.5 power) periods (see Figure A.13). The final observing season is separated from the earlier observations by >750 d, and it shows a very well sampled, prominent outburst, whose decay is followed by a re-brightening, which is exaggerated by the de-trending of the light curve (because the gradient of the fitted trend seems to have been too negative), and contains a single high-flux observation ~50 d after the outburst. Since the outburst also contributes to the dominant 137.3 d signal, this combination of effects resulted in a large artefact detection cluster in the WWZ.

4.4.0.8 SXP59.0

The WWZ of SXP59.0 contains two large detection clusters in the final observing season of OGLE-IV, with peaks at the periods of 48.6 d (76.9 power) and 19.4 d (67.3 power), which are



(a) (b)

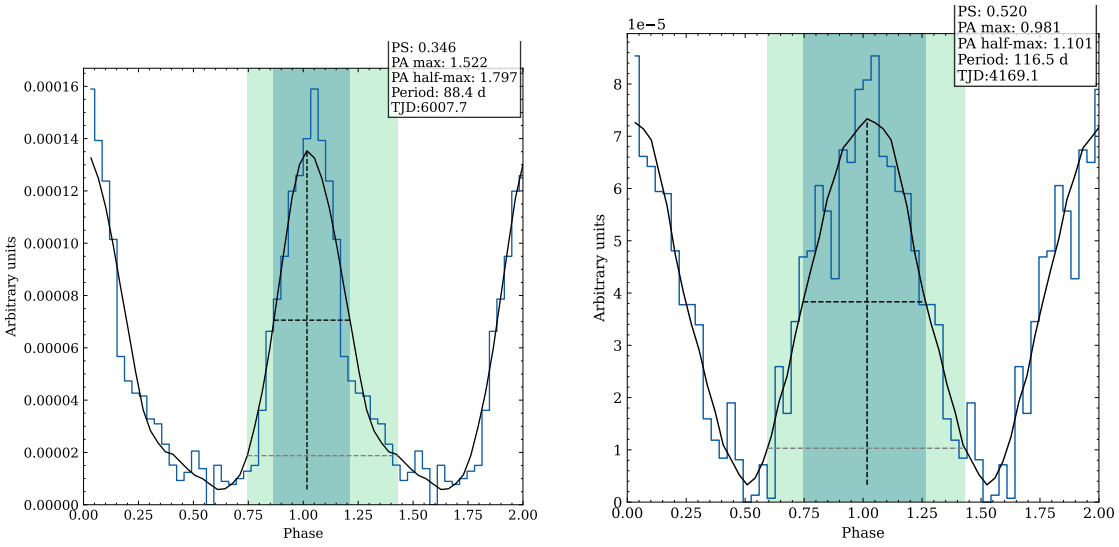
FIGURE 4.13: Phase-folded profiles of signals detected in the light curve of SXP59.0.

similar to the artefact found in the WWZ of SXP46.6. The light curve of that season shows three peaks in the time span of ~50 d.

Other prominent detections include a 249.7 d period (71.4 power, 1220.8-7446.3 TJD), which has a noisy, sinusoidal profile (PS=0.412, PA=1.430) and is present from the start of the light curve, but disappears for the last four seasons of OGLE-IV. Periods of 160.2 d (41.4 power, 5117.6-6632.3 TJD), and 147.5 d (34.2 power, 4857.4-6305.3 TJD), have similar profiles but are present in shorter time intervals. A period of 117.7 d (38.7 power) is detected between 3349.4-6452.1 TJD, which coincided with the time interval when the source faded from 15.2 mag to 15.8 mag and then re-brightened to 15.2 mag (see the light curve in Figure 4.4d). The 117.7 d period is the closest to the X-ray period of 122.10 ± 0.38 d reported by Galache et al. (2008) and the 122.0 d given by Coe and Kirk (2015), however the phase-folded profile of this modulation is wide and asymmetric, with a left skew (see Figure 4.13a), which does not support the possibility of orbital origin of the signal. Furthermore, Schmidtke et al. (2013) found no evidence of the 122 d period in the OGLE-III data.

Coe and Orosz (2000) found a period of 14.26 d in the OGLE-II data, but they found that folding the data on that period produced only a small sinusoidal modulation. In the WWZ transform, I find a signal with a period of 15.4 d (33.3 power, 1054.0-1220.8 TJD) in the second season of OGLE-II, which reappears in the seventh season of OGLE-III at 15.2 d (21.4 power, 4203.4-4283.5 TJD), with some power also present in the sixth season. It has a sinusoidal profile with PS of 0.492 and PA of 1.223 (see Figure 4.13b).

Additionally, Schmidtke and Cowley (2005) reported peaks in the periodogram at 20, 30 and 60 d periods in OGLE-II data, while Rajoelimanana et al. (2011) reported a period of 62.15 ± 0.04 d, however only the 20 d period was detected in the WWZ, with little power at the 30 and 60 d



(a) (b)

FIGURE 4.14: Phase-folded profiles of signals detected in the light curve of SXP91.1.

periods in the first OGLE-II season. A period of ~40 d is also detected in the second, seventh, and eighth seasons of OGLE-III, all showing sinusoidal profiles.

4.4.0.9 SXP74.7

The WWZ of SXP74.7 shows a signal with a period of ~33.3 d, detected in the OGLE-II and III data, as well as the final season of OGLE-IV, while the other seasons of OGLE-IV are likely sampled too sparsely to capture the modulation. The phase-folded light curve profiles have sinusoidal PS-PA values. The optical and X-ray periods of ~33.3 d were previously reported in the literature (Schmidtke and Cowley, 2005; Edge, 2005; Rajoelimanana et al., 2011; Bird et al., 2012; Schmidtke et al., 2013; Coe and Kirk, 2015).

4.4.0.10 SXP91.1

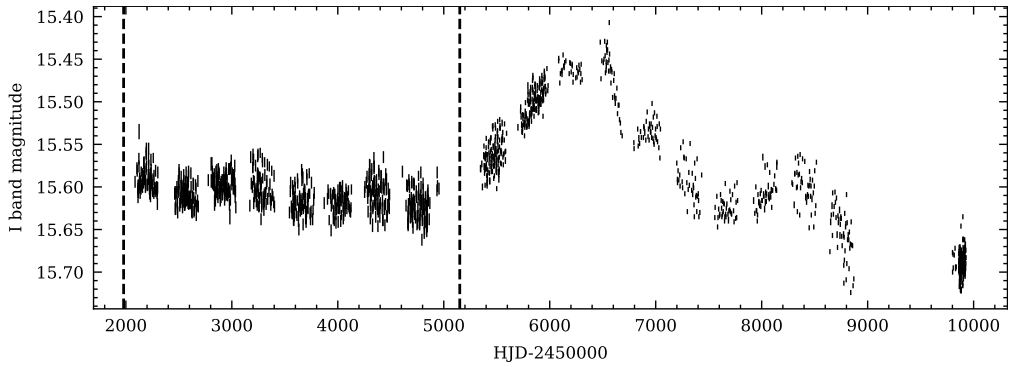
The WWZ of SXP91.1 shows a detection of a persistent signal with a period of 88.4 d (408.7 power), which corresponds to the period of the prominent oscillation of the light curve, and with the shape metrics of PS=0.346 and PA=1.522, it is a candidate for the orbital period modulation (see Figure 4.14a). Two other persistent signals are present in the WWZ of the source; one with a 116.5 d period (58.8 power) with its sinusoidal profile shown in Figure 4.14b, as well as one with a 71.2 d period, which was not detected by the pipeline however, likely due to the artefact detection cluster, which is present in the last seasons of OGLE-IV for this source as well. Galache et al. (2008) showed that SXP89.0 is in fact a misidentification of SXP91.1, and gave two X-ray periods, 87.6 ± 0.3 d for SXP89.0 and 117.8 ± 0.5 d for SXP91.1. Townsend et al. (2013) determined the orbital period measurement of 88.42 ± 0.14 d, while Kennea et al. (2018b) reported a 89.25 ± 2.54 d period in the X-ray data. The optical period of ~88 d was also

reported by Bird et al. (2012) and Schmidtke et al. (2004). The latter also reported a period of 6.2 d in the MACHO data of this source, but I find no evidence of it in the WWZ scalogram.

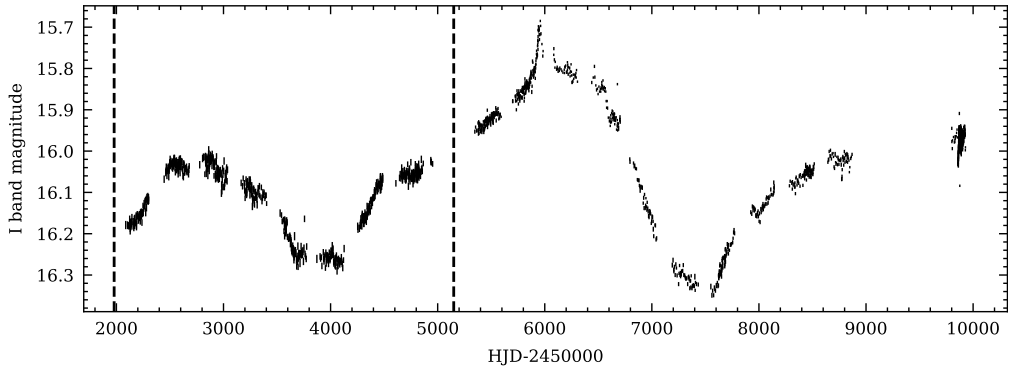
4.4.0.11 SXP101

The WWZ scalogram of SXP101 shows a strong signal with a period of 22.0 d, detected as three detection clusters in OGLE-III and the first two seasons of OGLE-IV; it is detected between 2086.9-2700.8 TJD with peak power of 49.7, between 3134.6-3881.9 TJD with peak power of 52.3, and between 4215.6-6117.3 TJD with peak power of 232.7. Starting from the third seasons of OGLE-IV the signal is not detected, but some power can still be found at the period of 22.0 d, as well as periods of 11.2, 7.3, 5.5 d, which are likely to be harmonics of the 22.0 d signal (see Figure 4.16.). As shown in Figure 4.15a, the source had a fairly flat light curve throughout OGLE-III, but in the first four seasons of OGLE-IV it brightened by ∼0.14 mag, reaching a maximum in the fourth season, and then fading by ∼0.08 mag below the OGLE-III level. The cessation of the 22.0 d period in the third season of OGLE-IV coincides with the source reaching its maximum brightness, but the sampling cadence of OGLE-IV is also significantly reduced during that time, which seems to be the more likely cause of the cessation of the signal, especially since its harmonics persist. In the final season of OGLE-IV the sampling cadence improves again, reaching the median frequency of more than one observation per night, however only the 7.3 d harmonic persists, and the 22.0 d period does not re-appear. Assuming orbital period origin of the 22.0 d signal, the cessation of the signal could be explained by the fact that the source reaches its minimum brightness in the last season of OGLE-IV, which can indicate that the decretion disc becomes too small to be disrupted by the neutron star. Galache et al. (2008) reported a maximum of the X-ray periodogram at 25.2 d, however, below the significance level. Furthermore, I find that the optical light curve phase-folded on the 22.0 d period has the sinusoidal profile, not expected for a signal of orbital period origin. However, Coe and Kirk (2015) gave the X-ray orbital period of 21.9 d.

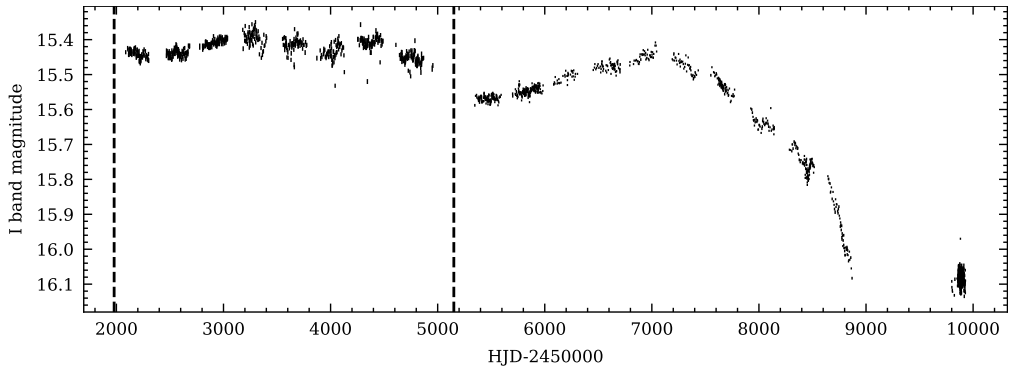
The WWZ also gives a detection of a signal with a period of 3.0 d (24.4 power, 2954.4-2987.8 TJD) in the third observing season of OGLE-III, with some power also present in the following seasons, until the first season of OGLE-IV. Phase-folded profile of the detection is sparse, with multiple gaps. Figure 4.17 shows a Lomb-Scargle periodogram of the third observing season of OGLE-III in the range of frequencies between $0.04 - 5\mathrm{d}^{-1}$. The most prominent peaks are found at the periods of 3.005, 1.508, 0.750, 0.600, 0.375 and 0.273 d, and they correspond to the alias pattern expected for an NRPs with the period of 0.750 d (the expected values are 3.000, 1.500, 0.600, 0.375, and 0.273 d, for the nominal sampling frequency of 1 d). Additionally, this source showed an NRP-like periodicity that preceded an increase in brightness, which could be an indication of a link between the NRP and the growing of the decretion disc.



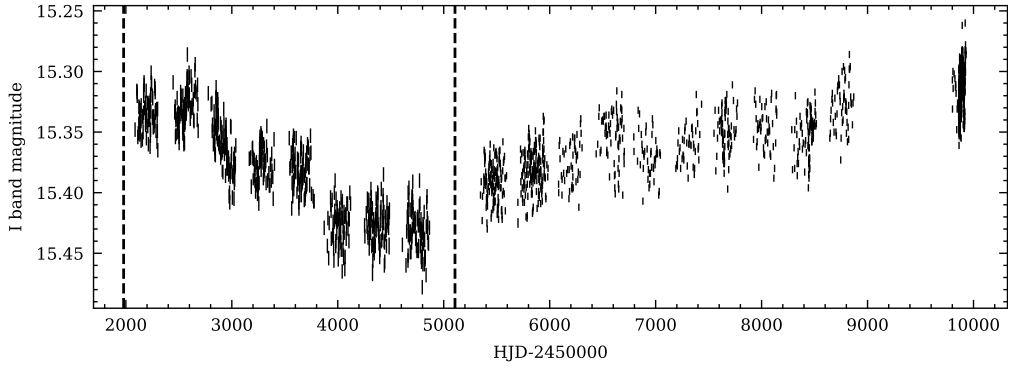
(a) SXP101



(b) SXP138



(c) SXP152.1



(d) SXP169.3

FIGURE 4.15: I-band OGLE light curves of analysed sources. OGLE-II, III and IV are separated by black dashed lines.

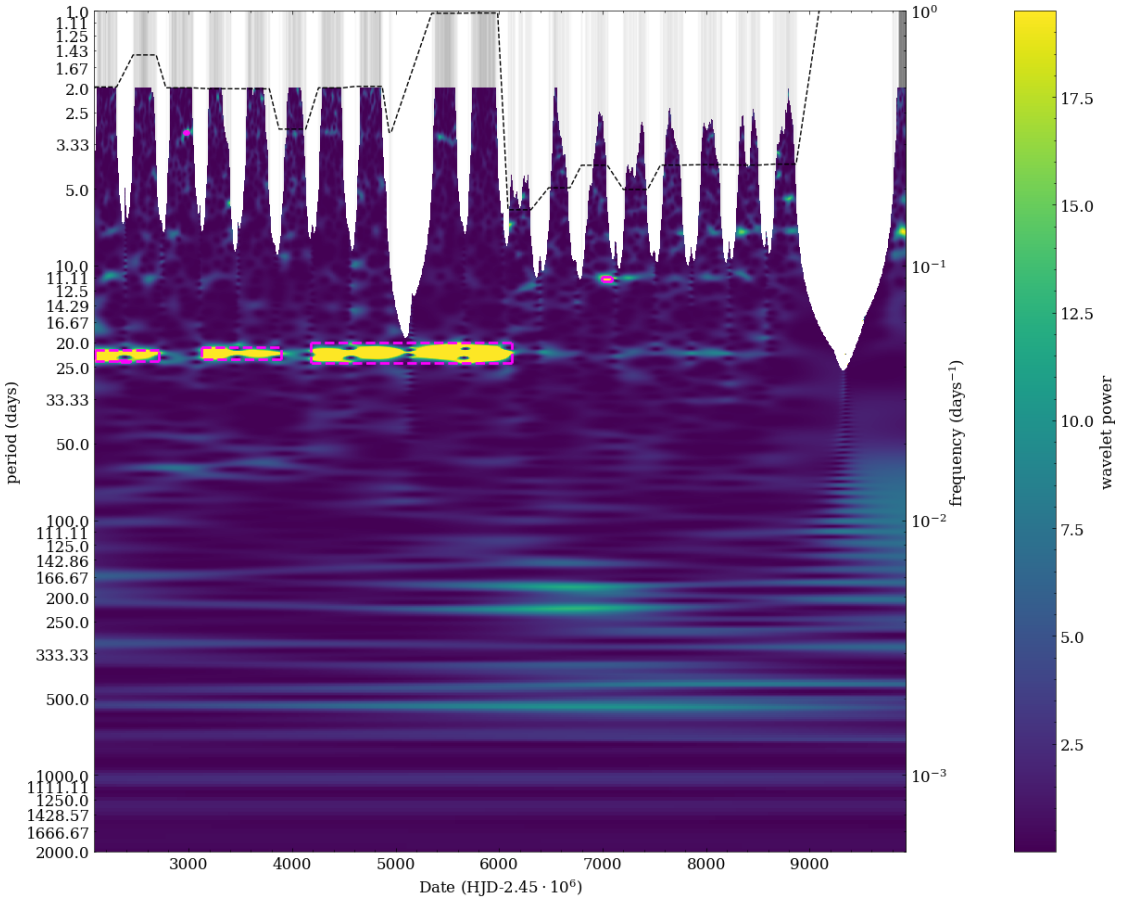


FIGURE 4.16: WWZ transform of SXP101 light curve. Grey vertical lines indicate the observation timestamps. The series of horizontal dashed black lines shows the quasi-Nyquist frequency, calculated as the median time difference between time stamps in every observing season. The wavelet power value is capped at 19.5 (corresponding to $\alpha = 0.05$ for an F-distribution with ∞ and 2 degrees of freedom). Dashed magenta boxes indicate the most extreme pixels which exceed the $\alpha = 0.05$ power level, for every detection.

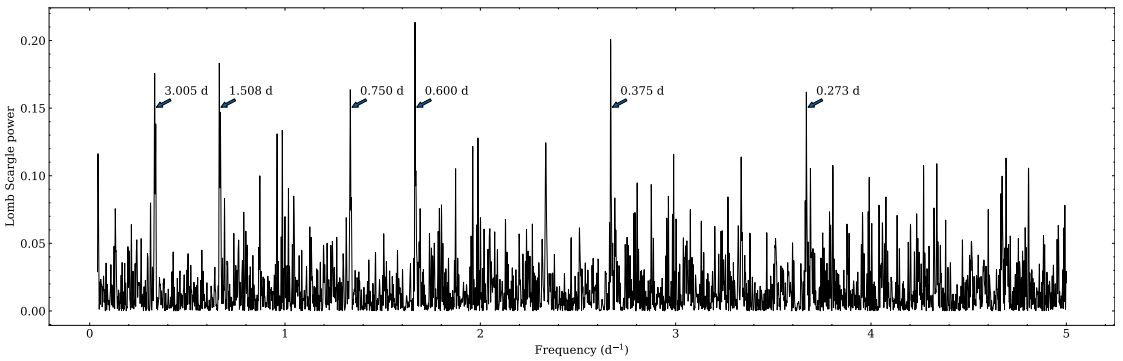


FIGURE 4.17: Lomb-Scargle periodogram of a section of SXP101 light curve between 2777-3038 TJD, over the range of frequencies expected for a non-radial pulsation of the Be star. The most prominent peaks are labelled with their corresponding periods.

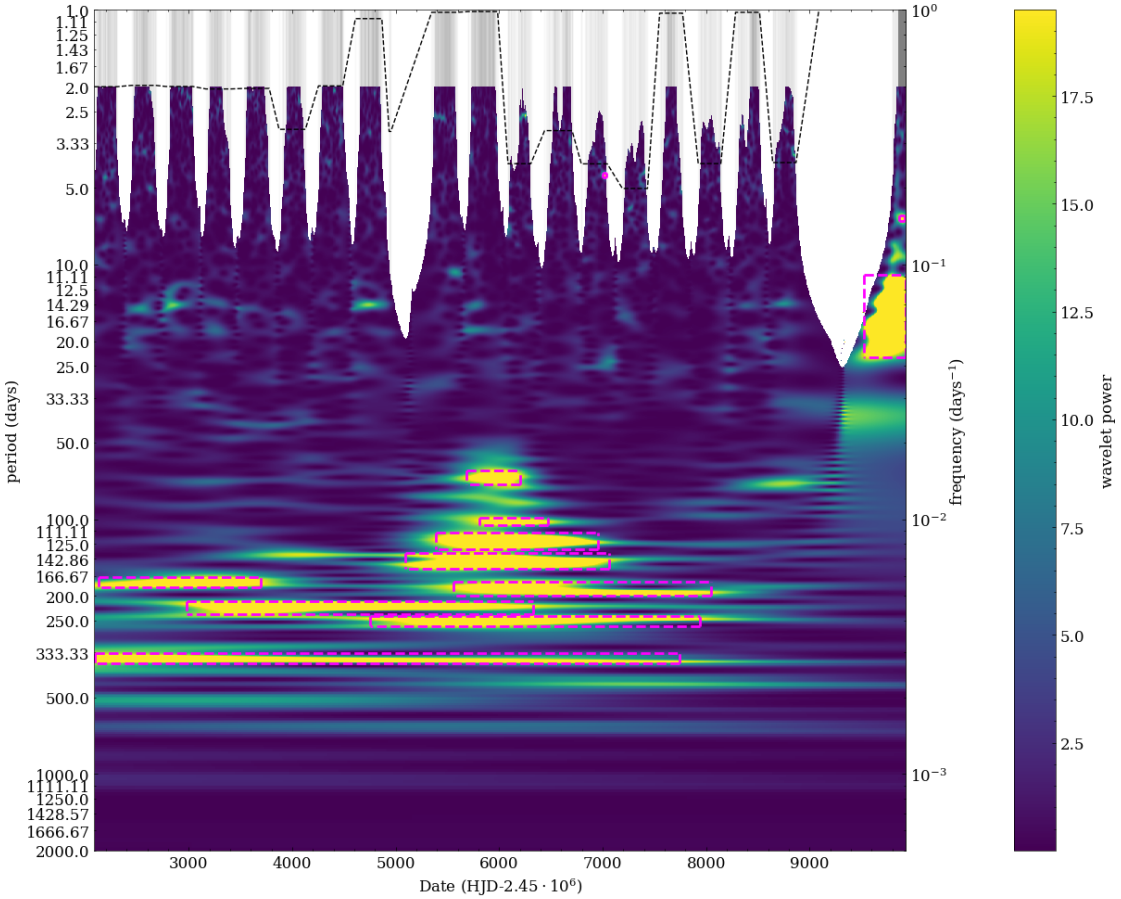
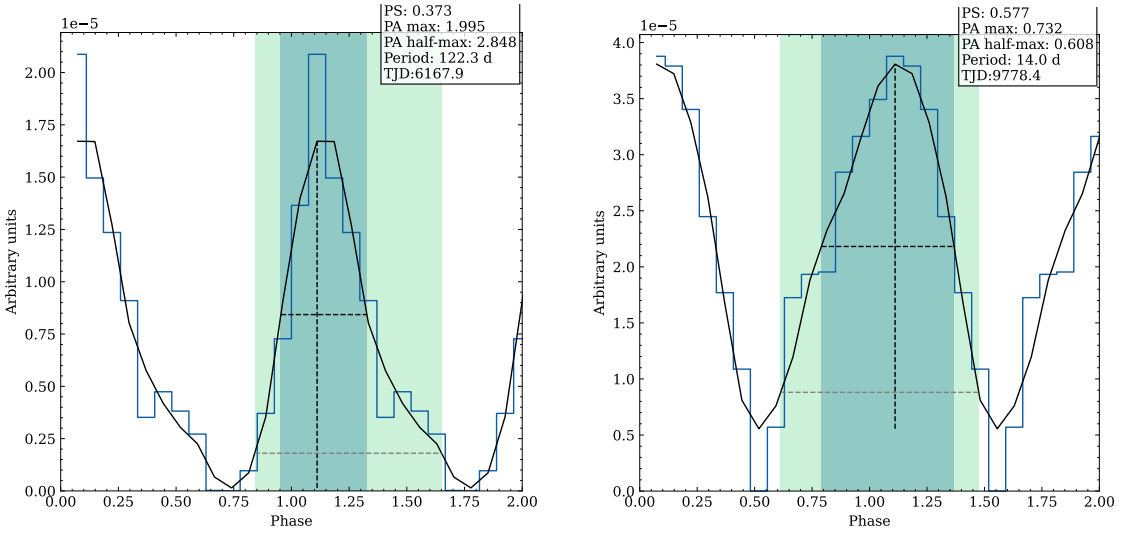


FIGURE 4.18: WWZ transform of SXP138 light curve. Grey vertical lines indicate the observation timestamps. The series of horizontal dashed black lines shows the quasi-Nyquist frequency, calculated as the median time difference between time stamps in every observing season. The wavelet power value is capped at 19.5 (corresponding to $\alpha = 0.05$ for an F-distribution with ∞ and 2 degrees of freedom). Dashed magenta boxes indicate the most extreme pixels which exceed the $\alpha = 0.05$ power level, for every detection.



(a) (b)

FIGURE 4.19: Phase-folded profiles of signals detected in the light curve of SXP138.

4.4.0.12 SXP138

The WWZ scalogram of SXP138 shows a detection with a period of 122.3 d (47.9 power, 5410.4-6925.4 TJD) which corresponds to the time interval when the brightness of the source exceeded $I = 16$ mag (see the light curve in Figure 4.15b). Edge et al. (2004) reported a similar period of 125 ± 1.5 d in MACHO object 207.16202.50, which is coincident with SXP138. Figure 4.19a shows that the phase-folded light curve of this WWZ detection has a FRED-like profile. In the same time interval I also find a 102.2 d periodicity (24.8 power, 5817.5-6464.9 TJD), which is close to the weak X-ray period of 103.6 d reported by Galache et al. (2008). The phase-folded profile shows some asymmetry, but it is still in the sinusoidal region on the PS-PA plane (PS=0.435, PA=1.984). Coe and Kirk (2015) gave the X-ray orbital period of 125.0 d.

Schmidtke et al. (2013) report an NRP with a period of 0.933 d in the second, third, and eight observing seasons of OGLE-III. In those seasons, the WWZ shows some power at the period of ∼14 d, see Figure 4.18. I also find a strong detection in the last season of OGLE-IV, with the period of 14.0 d (255.8 power, 9628.2-9928.5 TJD). It has a sinusoidal profile (see Figure 4.19b), and it could be and alias of the 0.933 d NRP, which is expected at the 13.925 d period for a 0.933 d signal with a sampling frequency of 1 d. The power at the ∼14 d period seems to be present in the observing seasons, with the median observation frequency of $\geq 0.5 \mathrm{d}^{-1}$ and $15.92 < I < 16.06$ mag.

The WWZ also returns detections at the periods of 4.4 d (42.6 power, 7005.4-7025.5 TJD, PS=0.396, PA=1.475), 6.6 d (22.4 power, 9875.2-9915.2 TJD, PS=0.593, PA=1.405), 148.7 d (42.5 power, 5156.8-6952.1 TJD, PS=0.342, PA=1.486), as well as several other periods in the range 180-361 d with noisy phase-folded profiles, which do not have well-defined shapes.

4.4.0.13 SXP152.1

The WWZ scalogram of SXP152.1 shows two strong detections of a ∼37 d periodicity with a sinusoidal profile; in the first two seasons of OGLE-IV with a 36.4 d period (43.0 power, 5123.5-5830.9 TJD, PS=0.575, PA=1.260), and in the final season with a 37.8 d period (72.2 power, 9294.5-9928.5 TJD, PS=0.594, PA=0.907). The period was reported by Schmidtke et al. (2013) at 36.28 d and interpreted as an alias of an NRP with a period of 1.026 d.

The WWZ also detects a sinusoidal 72.1 d periodicity (26.1 power, 2173.7-2654.2 TJD, PS=0.620, PA(alt.)=0.915) in the first two seasons of OGLE-III, and a 190.7 d period (21.9 power, 3061.3-4115.7 TJD), which shows peak power in the fifth season, but seems to be present throughout the OGLE-III data. The latter period has a noisy, fairly narrow, left-skewed profile (PS=0.359, PA=0.395). Similar periods re-appear in the ninth and tenth seasons of OGLE-IV, this time at 80.1 d (29.6 power, 8200.0-9281.2 TJD, PS=0.488, PA=0.506) and 188.9 d (31.3 power, 8253.4-9428.0 TJD, PS=0.565, PA=2.442), with the longer period showing an asymmetric profile again (see Figure 4.20a).

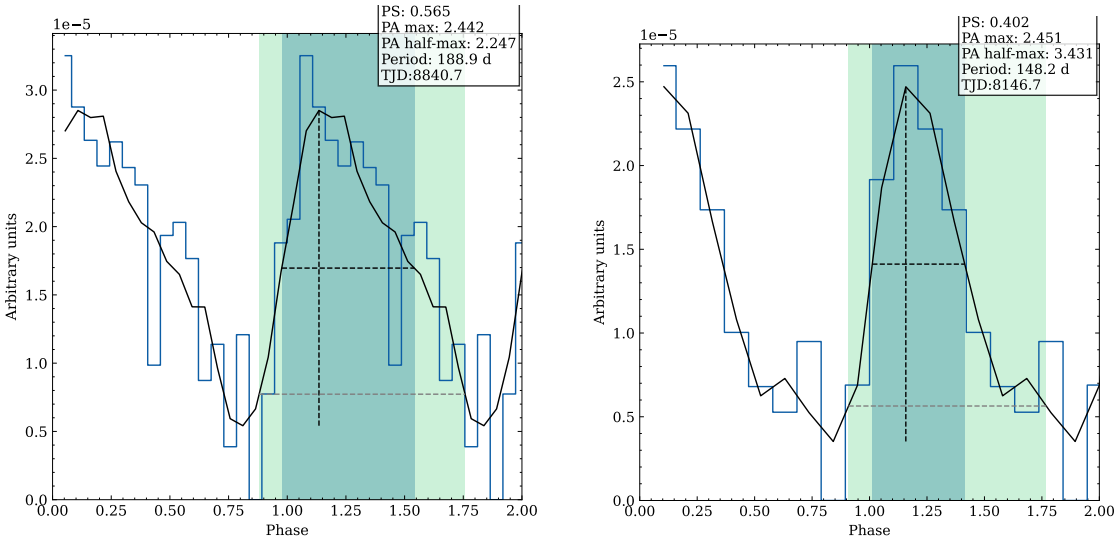


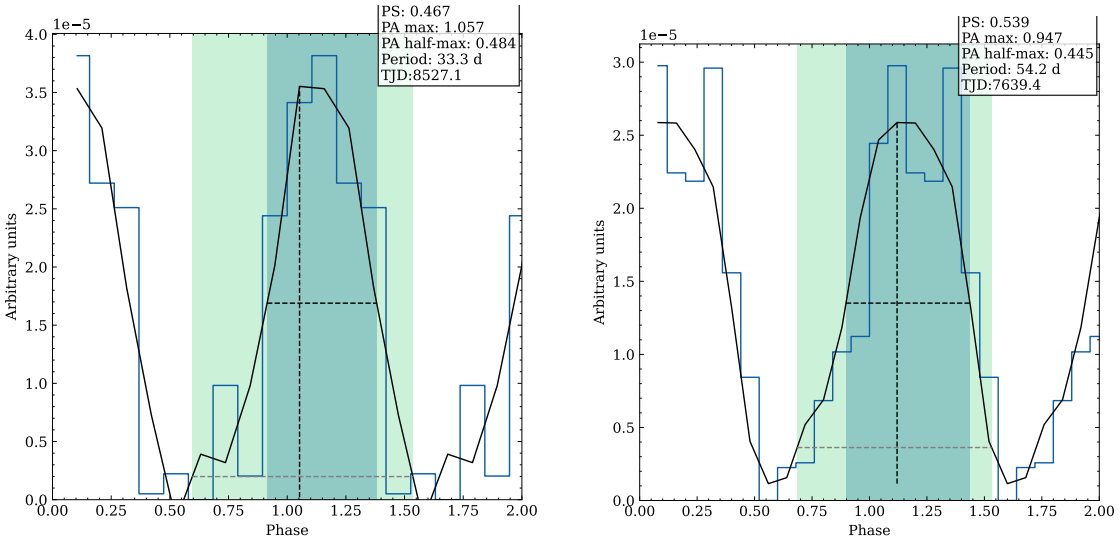
FIGURE 4.20: Phase-folded profiles of signals detected in the light curve of SXP152.1.

Other periods are also detected in the same time interval, including a 148.2 d period (23.5 power, 7592.7-8700.6 TJD) with a FRED-like profile (see Figure 4.20b), which also shows some power in earlier seasons of OGLE-III and IV, possibly as a doublet of peaks with another period of ∼125 d. In the tenth seasons of OGLE-IV, the 148.2 d period is dominated by the 134.9 d period (22.6 power, 8713.9-9154.4 TJD, PS: 0.397, PA=2.563), which also has a FRED-like profile, similar to the one of the 148.2 d modulation. The appearance of the signals in the latter half of the OGLE-IV data coincides with the rapid fading of the source, which becomes dimmer by $\Delta I \sim 0.6$ mag (see Figure 4.15c). The steep gradient of the de-trending model caused some significant changes to the shape of the light curve, which makes the interpretation of the results more difficult. Other periods detected in the final seasons of OGLE-IV include 374.8 d (25.3 power, 7686.2-9928.5 TJD), 267.4 d (21.7 power, 7118.9-7946.4 TJD), 218.3 d (34.8 power, 7699.5-9468.1 TJD), however they have noisy profiles and are likely caused by the gaps in the data.

4.4.0.14 SXP169.3

The WWZ scalogram of SXP169.3 shows detections of 68.4 d (32.9 power, 2287.1-3081.3 TJD, PS=0.330, PA=2.589) and 228.5 d (24.2 power, 2086.9-3521.8 TJD) periods, where the former has a FRED-like profile and corresponds to the orbital period, whilst the former has a noisy profile and is likely caused by the window function of the data. The optical period of ∼68 d was previously reported by Schmidtke et al. (2006); Bird et al. (2012); Schmidtke et al. (2013), whilst Galache et al. (2008) reported an X-ray period of 68.54 ± 0.15 d.

Several sinusoidal oscillations with periods in the range of 33-55 d are detected throughout OGLE-IV (see Figure 4.21 for examples showing the best definition of phase-folded profiles).



(a) (b)

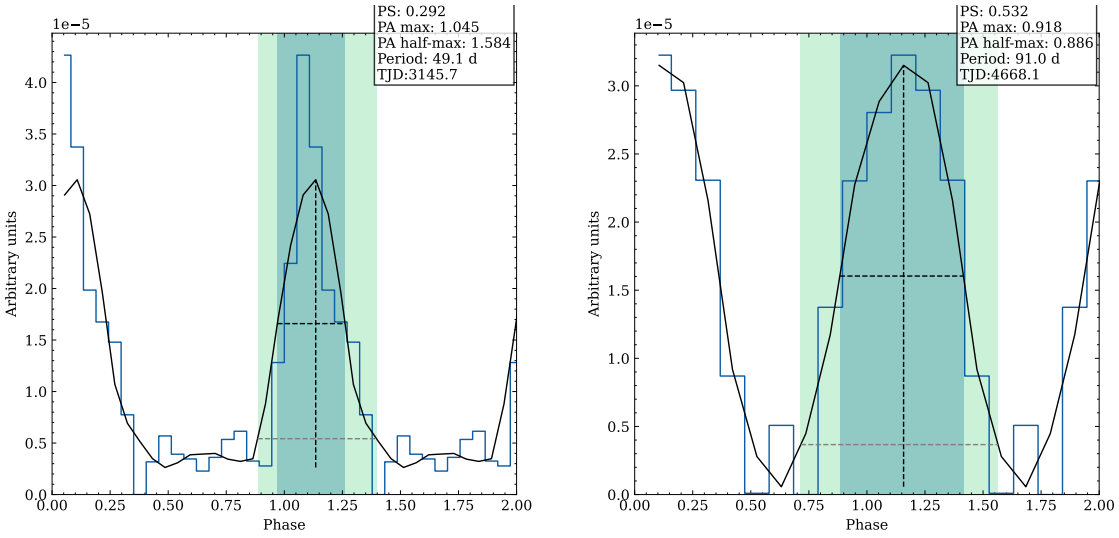
FIGURE 4.21: Phase-folded profiles of signals detected in the light curve of SXP169.3.

A period with a more asymmetric, FRED-like profile is detected at 34.6 d (21.8 power, 7459.3-7652.8 TJD, PS=0.441, PA(alt.)=2.136), and is a potential harmonic of the 68.4 d period.

Schmidtke et al. (2013) found an NRP with a period of 0.755 d in the OGLE-III data, and it is likely to be the cause of the power at the ∼3 d period, which is present in the WWZ of both OGLE-III and IV. This power only exceeds the detection threshold in two observing seasons, first at the period of 3.3 d (20.0 power, 6952.1-6958.7 TJD) and then at 4.0 d (76.1 power, 9828.4-9915.2 TJD). In the first season of OGLE-III, this power is also closer to the 4.0 d period, then drifting to ∼3 d period until the very last season of OGLE-IV. There might be anti-correlation between the period of this NRP alias and the brightness of the source, which reaches maxima in the first, second, and last observing seasons (see the light curve in Figure 4.15d).

4.4.0.15 SXP264

The WWZ transform of the SXP264 light curve shows a strong detection of a FRED-like, 49.1 d period (77.5 power, 1930.5-4361.0 TJD), which lasts through most of the OGLE-III phase, however it fades in the last two seasons, where a sinusoidal oscillation is detected, first at a 91.0 d period (71.6 power, 4020.4-5315.8 TJD), and then at a 75.7 d period (65.4 power, 4227.4-5369.2 TJD, PS=0.484, PA=1.300). Phase folded profiles of the 49.1 d and 91.0 d periods are shown in Figure 4.22. No strong periods are detected in the OGLE-IV data of the source, likely due to the sparse sampling (see the light curve in Figure 4.23a), however the WWZ shows some power around the ∼50 d period, starting from 5800 TJD. The 49.1 d signal corresponds to the orbital period of 49.2 d (Coe and Kirk, 2015).



(a) (b)

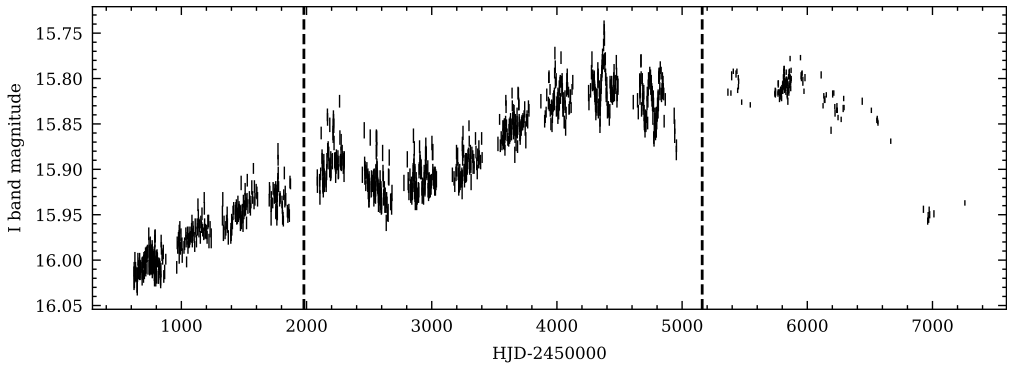
FIGURE 4.22: Phase-folded profiles of signals detected in the light curve of SXP264.

4.4.0.16 SXP293

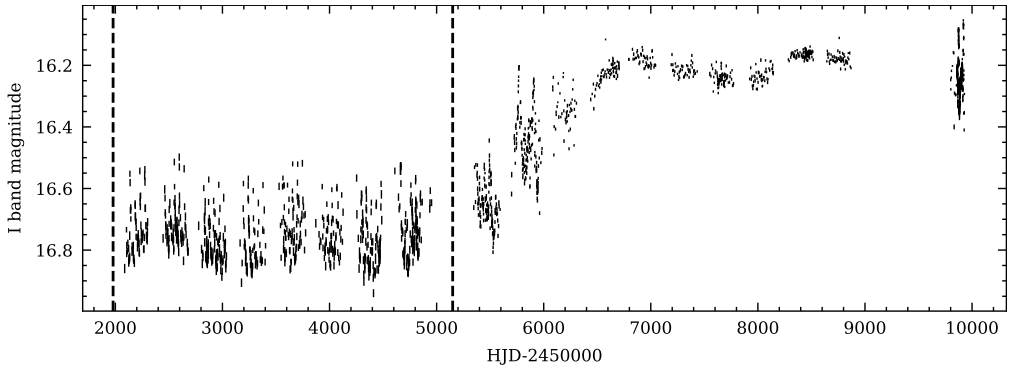
The WWZ of SXP293 shows a strong detection of a 59.7 d period (419.6 power), which persists throughout the light curve. Figure 4.24a shows the phase-folded light curve, which has a clear FRED-like profile. A dip in the profile right before the maximum was reported by Schmidtke et al. (2013), and it is also visible in the figure. This modulation is caused by the orbital period; Kennea et al. (2018b) reported a 59.62 ± 0.95 d period in the X-ray data.

4.4.0.17 SXP304

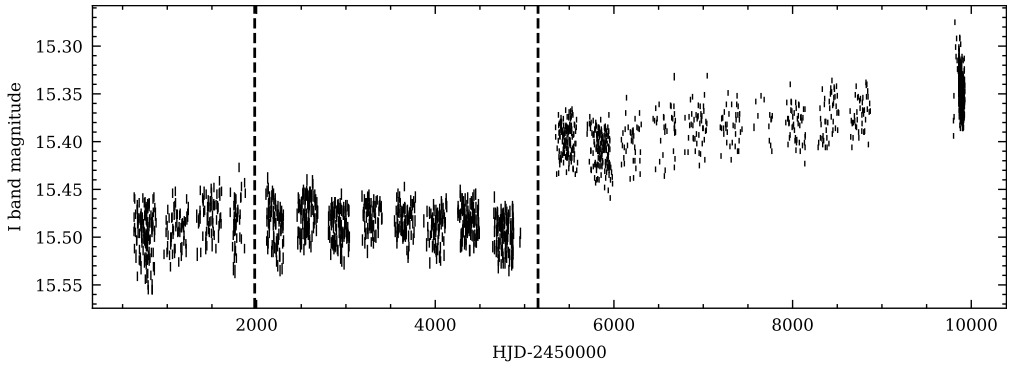
Three signals are detected in the WWZ spectrogram of SXP304; they have periods of 341.3 d (47.9 power, 627.9-7560.0 TJD), 171.4 d (47.7 power, 2749.6-5531.8 TJD) and 390.0 d (21.6 power, 3783.7-5411.7 TJD). I find no evidence of the 520 ± 12 d period reported by Schmidtke et al. (2006). The 171.4 d and 390.0 d periods have sinusoidal profiles, whilst the 341.3 d period has a narrow peak with an asymmetric profile, shown in Figure 4.24b. The 341.3 d period persists throughout the light curve, but its power drops below the detection threshold during the last four seasons of OGLE-IV. Kennea et al. (2018a) reported the detection of two X-ray outbursts from this source, first one between 57736.97-57764.57 MJD, and the second one starting from 58407.54 MJD (670.5 d after the start of the first outburst), and peaking at 58442.34 MJD, with no other X-ray detections outside those times. Based on these detections, they estimated the candidate orbital period of ~662 d, which is close to the double of the 341.3 d optical period.



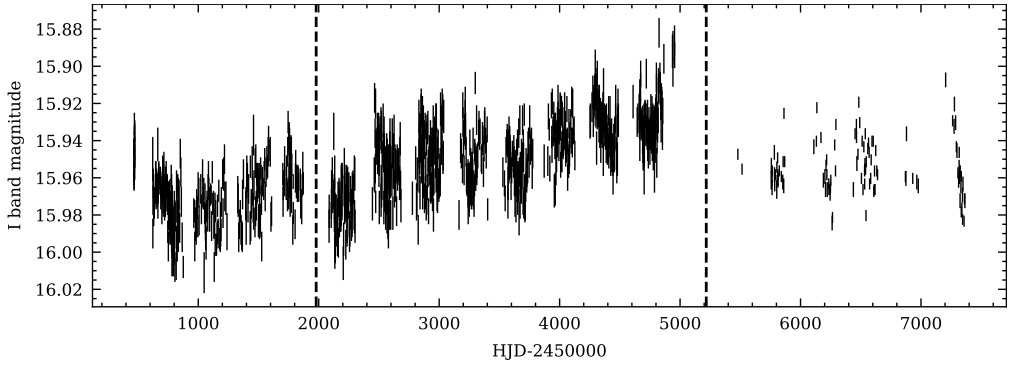
(a) SXP264



(b) SXP327

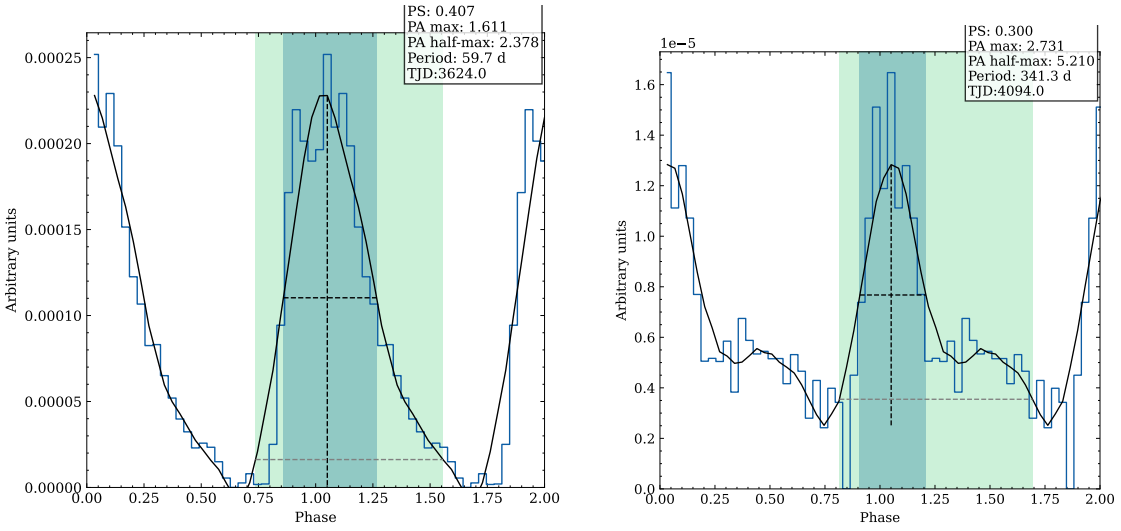


(c) SXP726



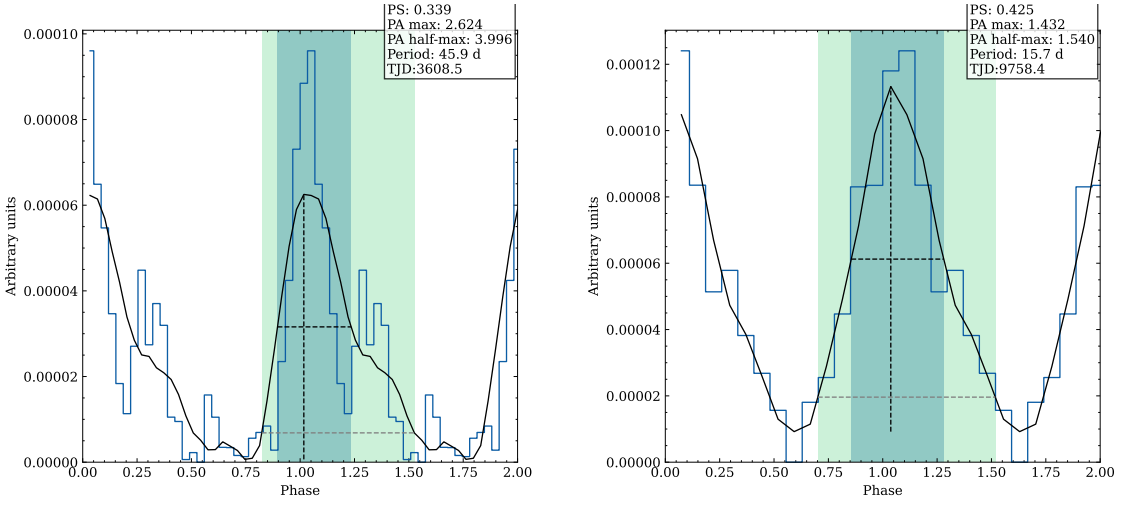
(d) SXP893

FIGURE 4.23: I-band OGLE light curves of analysed sources. OGLE-II, III and IV are separated by black dashed lines.



(a) SXP293 (b) SXP304

FIGURE 4.24: Phase-folded profiles of signals detected in the light curves of SXP293 and SXP304.

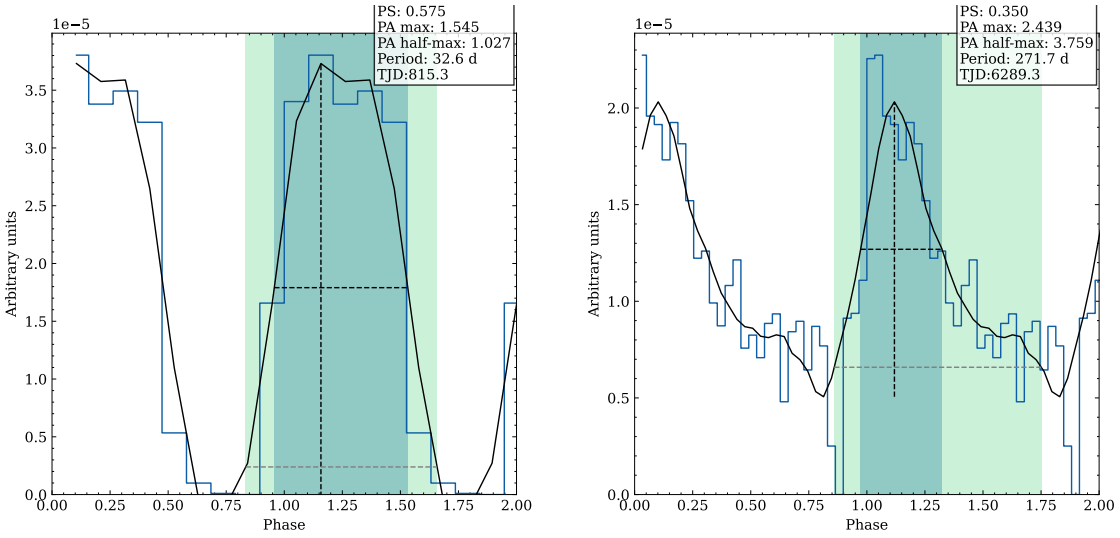


(a) (b)

FIGURE 4.25: Phase-folded profiles of signals detected in the light curve of SXP327.

4.4.0.18 SXP323

The WWZ of SXP323 showed a 964.1 d period (28.9 power, 621.8-7079.8 TJD) with a noisy, sinusoidal profile (PS=0.531, PA=1.499), which persists throughout the light curve, but its power falls under the detection threshold in the last six seasons of OGLE-IV. However, there is a very strong 2.4 d signal detected in every season where the sampling was good enough to test for this period. The signal has a well-defined sinusoidal profile, and it is an alias of an NRP reported by Schmidtke et al. (2013), with the period of the pulsation drifting from 0.7079 to 0.7086 d.



(a) (b)

FIGURE 4.26: Phase-folded profiles of signals detected in the light curve of SXP504.

4.4.0.19 SXP327

The WWZ transform of SXP327 shows a FRED-like periodicity with a double peak and the period of 45.9 d (104.9 power, 2086.9-5130.1 TJD) (see Figure 4.25a). This behaviour was previously reported by Coe et al. (2008), who also showed that it is caused by the 45.99 d orbital period.

In the rising phase of the light curve (shown in Figure 4.23b), during the first four seasons of OGLE-IV, the orbital modulation is dominated by detections of longer periods, in the range between 67-156 d. They are noisy, some are roughly sinusoidal, but others show a narrow central peak. These periodicities are likely caused by the fact that the large orbital oscillation was slightly deformed during de-trending. Longer periodicities which persist throughout the light curve have very noisy profiles and are likely caused by the sampling structure of the light curve. The final season of OGLE-IV shows a multi-frequency cluster of detections, with the most prominent period at 15.7 d (207.1 power), which has a sinusoidal profile, shown in Figure 4.25b. Other peaks in that cluster correspond to periods of 11.6 d (74.4 power, PS=0.564, PA=0.646), 24.5 d (57.5 power, PS=0.332, PA=0.460), and 48.3 d (48.2 power, PS=0.767, PA=0.263), with the last two corresponding to the orbital period and its harmonic.

4.4.0.20 SXP504

Three periods are detected in OGLE-II data of SXP504; a sinusoidal oscillation at 32.6 d (55.7 power, 621.8-1008.8 TJD, PS=0.575, PA=1.545) (shown in Figure 4.26a), as well as two asymmetric ones at 138.4 d (22.1 power, 621.8-1309.0 TJD, PS=0.308, PA=2.173) and 226.7 d (23.8 power, 621.8-1722.6 TJD, PS=0.534, PA=3.194). The 138.4 d asymmetric period might be a harmonic of the 271.7 d (47.7 power, 3684.1-8894.5 TJD) period, which shows some power

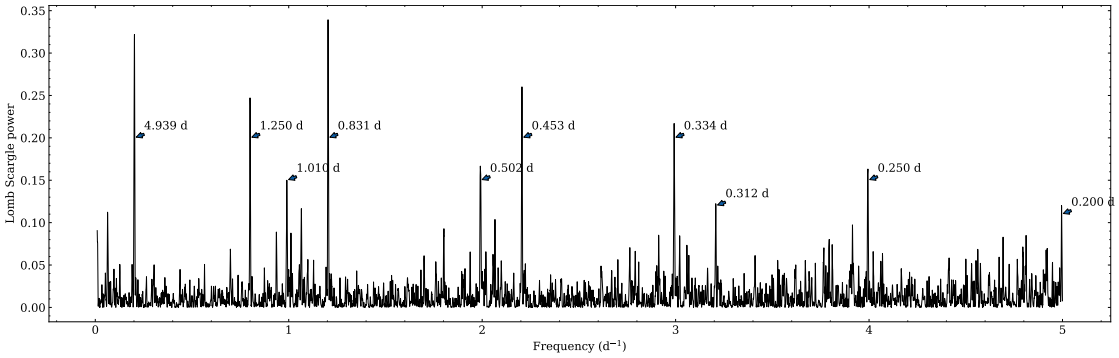
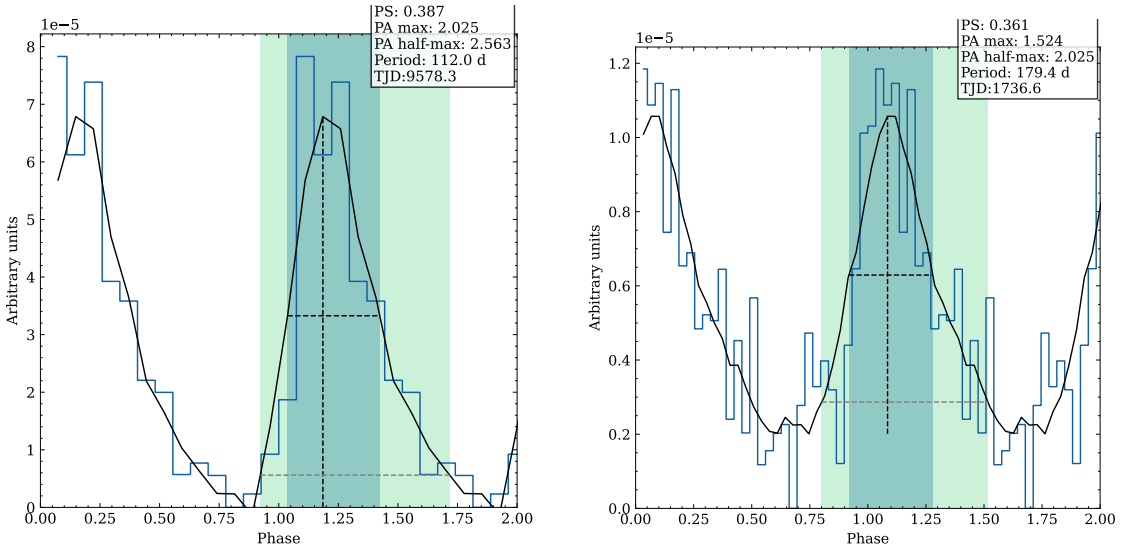


FIGURE 4.27: Lomb-Scargle periodogram of a section of SXP565 light curve between 628.9−998.8 TJD, over the range of frequencies expected for a non-radial pulsation of the Be star. The most prominent peaks are labelled with their corresponding periods.

throughout the entire light curve, but only exceeds the detection threshold starting from the fifth season of OGLE-III. Light curve phase-folded on the 271.7 d period is shown in Figure 4.26b. The 271.7 d period is most likely the orbital period; an X-ray period of 265.3 ± 2.9 d was reported by Galache et al. (2008), while Edge et al. (2005) found an optical period of 268.6 ± 0.1 d and a coinciding X-ray period.

4.4.0.21 SXP565

The WWZ transform of the SXP565 light curve shows the most prominent detection at the period of 86.1 d (74.7 power, 6997.8-9928.5 TJD), however this seems to be a component of a multi-frequency cluster of detection, found in the second half of the OGLE-IV data; the phase-folded light curve for this detection is noisy and lacks the structure of a real periodicity. Sinusoidal oscillations with periods of 80.1 d (74.2 power, 5055.1-6223.4 TJD, PS=0.594, PA(alt.)=1.039) and 64.6 d (63.8 power, 5041.7-5896.3 TJD, PS=0.555, PA=0.806) are detected in the first four seasons of OGLE-IV. The 80.1 d period seems to show a drift to a longer period of 91.0 d (54.6 power, 7418.4-9928.5 TJD, PS=0.423, PA=1.018) (see the WWZ in Figure A.25). A sinusoidal oscillation with a period of 152.3 d (32.3 power, 1977.5-3419.5 TJD, PS=0.457, PA=1.217) is detected in the first four seasons of OGLE-III, which corresponds to the previously reported orbital period (Galache et al., 2008; Bird et al., 2012; Coe and Kirk, 2015). Multiple other periods are also detected during limited time intervals, and they have sinusoidal profiles or undefined, noisy profiles. A sinusoidal oscillation with the period of 4.9 d (22.4 power, 769.1-809.2 TJD, PS=0.394, PA=1.037) is also detected in the first season of OGLE-II, with some power also present in the final season of OGLE-IV. The Lomb-Scargle periodogram of that season is shown in Figure 4.27, and its most prominent peak corresponds to the period of 0.831 d, which is the likely cause of the 4.9 d period. Additionally, several peaks from the periodogram are also found in Figure 25 of Schmidtke et al. (2013), which suggests that the 0.334 d oscillation is also present.



(a) SXP726 (b) SXP893

FIGURE 4.28: Phase-folded profiles of signals detected in the light curves of SXP726 and SXP893.

4.4.0.22 SXP701

The WWZ transform of the SXP701 data shows detections of a ∼2.2 d signal throughout OGLE-II, III and IV phases of the light curve. The phase-folded data shows a sinusoidal profile. The Lomb-Scargle periodogram of the data shows evidence of the 0.28 d and 0.64 d periods reported by Bird et al. (2012) and Schmidtke et al. (2013). Both of these signals are expected to produce aliases at ∼2.3 d period, therefore the periodicity found in the WWZ transform is likely an alias of these NRPs. I do not find evidence of the 412 d binary orbital period modulation (Coe and Kirk, 2015).

4.4.0.23 SXP726

The WWZ transform of the SXP726 data shows detections of a ∼3.3 d signal throughout OGLE-II, III and IV phases of the light curve. Eger and Haberl (2008a) reported this period along with a ∼2.35 d in OGLE-II data (but little power is found at this period in the WWZ transform). However, later Schmidtke and Cowley (2008) showed that higher frequency aliases at 0.302 d and 0.673 d periods have higher power. Therefore, the ∼3.3 d and ∼2.35 d periods are most likely aliases of the 0.302 d and 0.673 d NRPs. A FRED-like modulation with a period of 112.0 d is found in the most recent season of OGLE-IV, the phase-folded light curve is shown in Figure 4.28a. An orbital period of 112.0 d is shorter than expected for a pulsar with a spin period of 726 s (the expected orbital period of 661.4 d is given by Equation (1.1)). However, recently Haberl et al. (2022) reported that the source showed an average spin-down rate of 4.3 s/yr over 17 years of X-ray observation. Assuming that the spin-down had been ongoing before the measurement of the 726 s period (Eger and Haberl, 2008a), the orbital period of 112.0 d

would be viable. Furthermore, the source reaches its maximum brightness in the latest observing season (see the light curve in Figure 4.23c), which could be an indication of the decretion disc growth, so the sudden onset of the binary orbital modulation could be justified. However, the latest season of OGLE-IV covers a time interval of only 127.7 d, showing only two maxima of the modulation. This means that the detection of the 112.0 d signal is likely spurious, and more data is required to confirm whether any periodicity has appeared in the light curve of SXP726.

4.4.0.24 SXP756

The light curve of SXP756 shows very prominent outbursts with a period of ~395.4 d, caused by the binary orbital modulation. The orbital origin is confirmed by X-ray observations showing a period of 389.9 ± 7.0 d (Galache et al., 2008). Bird et al. (2012) reported an optical period of 393.6 ± 1.2 d, whilst Schmidtke et al. (2013) found 393.1 ± 0.4 d. The WWZ scalogram shows a persistent detection of the 395.4 d period, together with a series of its harmonics, which are most prominent in the OGLE-III phase. Other periods with noisy phase-folded profiles are also detected in the WWZ, which are most likely caused by the sampling pattern of the light curve. A period of 11.3 d is detected in OGLE-II, III and the first season of OGLE-IV. It has a sinusoidal profile, and Schmidtke et al. (2013) interpret it as an alias of an NRP with a period of 0.917 d.

4.4.0.25 SXP893

The WWZ transform of SXP893 data shows detections of a 3.7 d period, which is present throughout OGLE-II and the first three seasons of OGLE-III. This periodicity has a sinusoidal profile, and it is interpreted as an alias of an NRP with a period of 0.788 d. The periodicity is detected until ~3000 TJD while the source is steadily brightening, and there is no evidence of the period reappearing after that, however the source continued to brighten throughout OGLE-II and III, reaching a maximum of ~15.89 mag on ~5000 TJD. The source fades back to ~15.95 mag in OGLE-IV (see the light curve in Figure 4.23d). While the initial brightening of the source could be driven by the NRP, and the subsequent fading could be caused by the cessation of the pulsation, it is difficult to conclude if such a causality actually exists.

The WWZ shows a detection of a 91.2 d period (21.0 power, 1153.6-1531.8 TJD) in the OGLE-II phase, which was reported by Schmidtke et al. (2004), who suggested that the binary has an orbital period of either ~91.5 or ~187. A stronger period of 179.4 d (24.3 power, 466.5-3006.7 TJD) is detected in the WWZ of OGLE-II and the first four seasons of OGLE-III, and it is the better candidate for the orbital period, given its FRED-like profile (shown in Figure 4.28b), while the profile of the 91.2 d period is very noisy and rather sinusoidal. Schmidtke et al. (2013) also found a weak ~182 d period, but concluded that more data is needed to confirm it. The detection of a 366.0 d period (43.2 power, 3139.0-6977.6 TJD) found in the WWZ of the second half of the light curve is likely to be caused by the annual data trains.

4.4.0.26 SXP1323

The WWZ of SXP1323 shows detections of previously reported optical periodicities (Schmidtke et al., 2006; Bird et al., 2012); the 26.2 d period is present throughout the OGLE-II, III and IV data, and the 7.9 d period is detected until the fourth season of OGLE-IV, but beyond that it fades below the detection threshold. Both modulations have sinusoidal profiles, which is unexpected for the 26.2 d period, since it is associated with the binary period; Kennea et al. (2018b) reported a 25.80 ± 0.43 d period in the X-ray data. This could be caused by low eccentricity of the orbit. The orbital period of this system is also very short, given the spin period of ~1323 s, especially given that it has been shown to be spinning up (Carpano et al., 2017). A sinusoidal oscillation with the period of 2.4 d (56.8 power, 9861.8-9895.2 TJD) is also detected in the final season of OGLE-IV, with some power also present in earlier seasons. Schmidtke et al. (2013) interpreted the 7.9 d and 2.4 d periods as aliases of NRPs with periods of 0.886 and 0.412 d.

4.4.0.27 Remaining sources

WWZ transforms of SXP0.92 and SXP34.08 produce no detections. The OGLE-XROM catalogue includes other sources, like: SXP11.5, SXP140, SXP15.3, SXP172, SXP18.3, SXP2.37, SXP202, SXP280.4, SXP3.34, SXP342, SXP348, SXP455A, SXP455B, SXP6.85, SXP6.88, SXP645, SXP65.8, SXP7.78, SXP7.92, SXP8.80, SXP82.4, SXP967. However, de-trending of their light curves did not give a result that would be good enough to continue with the WWZ transform.

4.5 Discussion and summary

In this chapter, I introduced a method of detection and characterisation of periodic signals, and demonstrated how it can be applied to the unevenly sampled light curves of highly variable sources. After tuning the parameters of the method using synthetic data with a realistic observing pattern, I applied it to a sample of 28 Be X-ray binaries, thus detecting and characterising several periodic signals for the first time. The resulting insights show that this wavelet-based periodicity detection pipeline is a useful tool that can facilitate the automation of the time-frequency analysis of photometric light curves, which show variability in the time scales of days, months, and years. In this section, I discuss these results and limitations of the method, and outline directions for further work that could refine it.

4.5.1 Periodicities in the OGLE-XROM sources

Application of the automatic detection pipeline to the OGLE-XROM data resulted in the discovery of six new candidate periodicities, which are listed in Table 4.1 alongside their respective

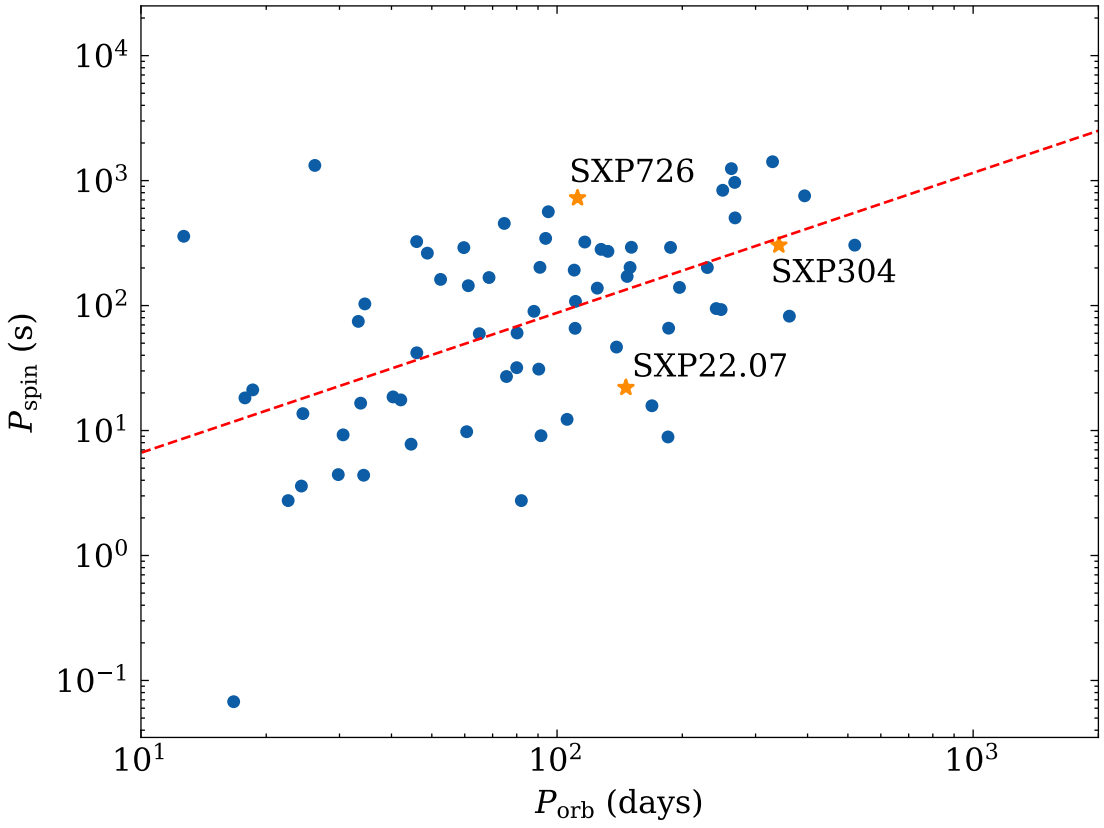


FIGURE 4.29: Corbet diagram of BeXRBs and candidates in the Galaxy and in the Small and Large Magellanic Clouds. Blue dots indicate sources from the catalogue by Raguzova and Popov (2005) (same as in Figure 1.3b). Orange stars indicate the sources whose FRED-like periodicity was identified using the method discussed in this chapter. The red, dashed line indicates the linear fit to the blue data points, as defined in Equation (1.1)).

sources and classifications. This analysis was supplemented with the manual fitting of Lomb-Scargle periodograms, in the few cases where sinusoidal modulations strongly suggested the presence of non-radial pulsations, which allowed me to identify the likely frequency of the pulsation resulting in the detected alias.

An oscillation with a 3.0 d period that was found in the first two seasons of OGLE-IV observations of SXP9.13 is likely to be an alias of an NRP with a 0.753 d period. A sinusoidal, 3.0 d periodicity was found in most of the OGLE-III data of SXP101, fading in the first season of OGLE-IV. The periodogram of these data showed that the 3.0 d signal could be an alias of an NRP with a 0.750 d period. Finally, the 4.9 d oscillation detected in the first season of SXP565 data is likely to be an alias of an NRP with a period of 0.831 d.

In three cases, the new candidate periodicities show FRED-like profile of the phase-folded light curve. These three periodicities are interpreted as candidate orbital periods of their corresponding sources and added to the Corbet diagram in Figure 4.29. The candidate orbital period of SXP304 is in good agreement with the value expected for a BeXRB with the spin period of 304 s, as given by Equation (1.1)). The candidate periods of SXP22.07 and SXP726 do not strictly agree

with expectation, but the periods place the sources within the scatter of the point cloud of the BeXRB sample in the spin-orbit plane, which means that they are not significant outliers.

SXP22.07 shows a signal with a 146.4 d period in the first two seasons of OGLE-III, when the source is at its brightest. However, 146.4 days would be a relatively long orbital period for a pulsar with a spin of 22.07 s (as the expected value is 29.2 d, according to Equation (1.1)), and I have not found evidence of spin-up of the neutron star in the literature. The WWZ transform of SXP726 shows a detection of a 112.0 d period in the last observing season of OGLE-IV, when the source reached its maximum brightness. As discussed in Section 4.4.0.23, the candidate orbital period of 112.0 d is shorter than expected for a BeXRB with a spin period of 726 seconds, but there is evidence for the spin-down of this neutron star. However, only one cycle of that candidate orbital period could be observed, so more data is required to verify whether the periodicity is real.

The WWZ scalogram of SXP304 shows a strong detection of a 341.3 d period with a FRED-like profile. Haberl et al. (2022) analysed eROSITA observations of SXP304 from November 2019 and found the spin period to be 302.29 ± 0.27 s, which is consistent with the value of 302.6 ± 0.4 s, measurements in November and December 2005 (Eger and Haberl, 2008b). This indicates that the spin period of the source over those 14 years showed little change. Yang et al. (2017) estimated the average spin period derivative to be $2.9 \times 10^{-4} \pm 8.7 \times 10^{-4}$ s day^{-1}. Assuming a constant spin period, the expected orbital period of a 302.29 s pulsar (as derived from Equation (1.1)) would be 302.5 d, which is another indication that the period of 341.3 d is a probable candidate for the binary orbital period. Alternatively, it could be a harmonic of the ~662 d period, which results from the observations of X-ray outbursts reported by Kennea et al. (2018a).

The time-frequency analysis and classification of periodicities was also an opportunity to probe the possible causality between the presence of NRPs and formation of the decretion disc. No conclusive evidence was found in the sample of analysed light curves, but some apparent correlations are note-worthy. For example, sinusoidal periods of ~15 d and ~20 d are found in the data of SXP59.0 in three separate time intervals, and in every one of them the source is fading. In the case of SXP101, the ~3 d alias of an NRP is present throughout most of the OGLE-III phase, when the brightness of the source is roughly constant. The source starts to brighten in the first season of OGLE-IV, and at the same time the alias shows a drift to lower frequency, ceasing after that. The scalogram of SXP169.3 shows a sinusoidal period of ~3 d throughout most of the light curve, but it was detected closer to the ~4 d period in the first and last observing seasons, around the times when the source reached the local brightness maxima. An NRP alias with a period of ~11.3 d is detected in the data of SXP756 until the first season of OGLE-IV, and that is when the smoothly increasing trend of the source brightness becomes more chaotic, with an overall decreasing trend. SXP893 shows an alias of an NRP at the period of ~3.7 d that lasts throughout OGLE-II and the first three seasons of OGLE-III data, while the source increases in brightness at a constant rate until the end of OGLE-III, when it starts to fade.

While these behaviours could be interpreted as potential correlation (or anti-correlation) between the presence of NRPs and changes in the size of the decretion disc, some sources showing persistent NRPs do not show consistent changes in brightness. For example, the scalogram of SXP701 shows a persistent detection of an NRP alias at a ∼2.3 d period, while the light curve shows a roughly constant brightness, following an interval of slow fading by ∼0.025 mag over the first eight seasons. In SXP726, an NRP alias with a ∼3.3 d period is detected throughout the light curve, while the source stays at a constant brightness level through OGLE-II and III, only starting to brighten in OGLE-IV.

4.5.2 Wavelet-based detection pipeline

The first transformation applied to the data as part of the pipeline is the de-trending step, where the trend component of the Prophet model is removed from the light curve. The Prophet model is very flexible, and gives good fits to light curves showing a large variety of behaviours. However, this flexibility means that it can easily over-fit or under-fit the data if not parameterised correctly. I find that the settings described in Section 4.2 are adequate for the majority of OGLE-XROM light curves, however a considerable number of sources (23 out of 51) could not be used in the remainder of the pipeline because the quality of their trend fit was not good enough. For example, light curves of some sources include large and rapid outbursts, which skew the trend component of the Prophet model and cause the surrounding observations to be affected disproportionately by de-trending. Other light curves show gradients which are changing too quickly for the model. In future work, a more robust de-trending method must be implemented. This could potentially be achieved by multi-steps de-trending where outliers are removed prior to the fitting of the Prophet model, for example by a sigma-clipping moving window.

The next step of the pipeline is the WWZ transform of the de-trended light curve. As explained in Section 4.3.2, estimating the confidence of WWZ detections is not always straightforward. The use of the F-distribution critical value to set the minimum power threshold is a way to avoid costly Monte Carlo simulations, which are often used to find the false positive rate. However, in case of the WWZ transform, the main challenge is caused by spurious, peaks with very high power that appear in the scalograms of poorly sampled light curves. The use of a relatively small wavelet decay constant and a high threshold for the effective number of data points help to mitigate the issue, but the small decay constant also causes the temporal resolution of the method to deteriorate. One potential way of addressing this could involve a modification to the wavelet function that would make the decay constant frequency-dependent, so that higher frequency wavelets would decay more slowly and the lower frequency ones would decay faster. This would result in a higher sensitivity to high frequency signals and a better time resolution of the low frequency signals. However, this would result in the method becoming more similar to a dynamic Lomb-Scargle spectrogram, and would result in the temporal resolution of the method at the high-frequency rate to deteriorate. In this context, the optimal rate of wavelet decay is

highly dependent on the sampling pattern of the data and the range of signal frequencies that we are interested in probing, so the decay would need to be tuned with these considerations in mind.

Once the scalogram is calculated and a detection threshold is set, the regions of the scalogram where signals are detected are identified through clustering. The currently implemented algorithm is greedy, and merges all the adjacent pixels with high power into a single cluster. This works well for most periodicities, but it becomes an issue when the scalogram shows a complex pattern of detections, including drifts in frequency. A different method of identifying detection clusters would be beneficial in such cases (see the scalogram of SXP565 is Figure A.25 for an example of a complex pattern). For example, fitting of a Gaussian mixture model could allow for the clustering of any arbitrary detection pattern, including drifts. If a Gaussian component of a cluster had the covariance of time and frequency dimensions close to zero, then this would indicate a stable frequency, resulting in the phase-folded analysis being triggered. However, a significant correlation or anti-correlation would result in the pipeline flagging the detection as a potentially drifting period, which would need to be analysed separately.

Following the identification of detection clusters and dominant frequencies within them, the next step of the pipeline involves the phase-folding of the light curve. Another limitation of the pipeline is the fact that a significant number of detections with periods >100 d have noisy profiles which resemble neither sinusoidal nor FRED-like behaviour, and this only becomes transparent during the phase-folding step of the pipeline. These detections are caused by the window function of the data, and ideally should be filtered out automatically. A possible criterion for this filter could be based on testing whether the phase-folded profile varies smoothly. For example, if the Savitzky-Golay fit to the folded profile does not raise and fall monotonically from the maximum, the detection should be flagged as anomalous or unreliable. This issue can also be observed for some high frequency periods, which are sampled sparsely and show empty bins.

The phase-folded light curve profiles are characterised based on the phase span and phase asymmetry metrics (Bird et al., 2012). These values are a useful way to quantify the shape of the detected modulations, however in case of several profiles presented in Section 4.4, they did not capture the shape accurately, often because the number of phase bins in the folded light curve was not optimal or because of anomalous values in singular phase bins skewing the result. The implementation of the alternative phase asymmetry metric helped to mitigate these issues in several cases. A way to combine this alternative metric with the one proposed by Bird et al. (2012) should be found in order to refine the classification of periodicities based on their position in the PS-PA plane. Alternatively, profiles could be classified based on their position in a three-dimensional space defined by all three metrics.

A possible extension of the pipeline could involve automation of the search for non-radial pulsation periods. Fitting of a Lomb-Scargle periodogram over the high frequency range would be triggered when a sinusoidal profile was detected, in order to test if high power aliases of that period are present in the data. Another possible extension to the periodicity detection pipeline could involve the calculation of gradients in the photometry of the sources exhibiting NRPs. A

quantitative analysis of a larger sample of cases could then be performed in order to address the question of causality between NRPs and the decretion disc formation.

Chapter 5

Conclusions

In this thesis, I introduce new methodologies of variability analysis for light curve data, and discuss their application to several X-ray binary systems. I addressed cases of variability in the time scales ranging from seconds to decades, in optical and X-ray energy bands. These methodologies are a contribution to the analytic toolboxes for the cases of both evenly and unevenly sampled data.

5.1 Fast X-ray variability

Chapters 2 and 3 introduce a data-driven method of light curve feature extraction and aggregation using an LSTM-VAE neural network and a GMM, show its application to data containing a wide variety of X-ray variability, and validate the utility of resulting features by conducting a set of supervised classification experiments. Presented results show that the proposed method is capable of producing features which represent light curve data in a meaningful way.

During the classification of GRS 1915+105 light curves in chapter 2, I tested two variants of data preprocessing, with 1 s and 4 s time bins, showing that the light curves binned to 1 s produced a significantly better classification performance, with the mean of weighted F1 and accuracy performance scores of 0.878 ± 0.027 and, 0.894 ± 0.027 respectively. I also showed that the resulting light curve fingerprint representation can be used for the unsupervised exploration of the latent data manifold, and quantification of similarity between light curves. A comparison of fingerprints of κ and ω observations showed that they can be very similar, and possibly lie on a continuum, showing a smooth transition between the two, which is not captured well by the Belloni et al. system of variability classification. In chapter 3, I explored the application of the fingerprint representation to the light curves of IGR J17091−3624, and found that the extension of the methodology to other sources will require modifications, particularly in the handling of differences in intensity between the sources. However, I found that similar light curves from the two sources were not assigned to the same class in a series of classification experiments, despite

an overlap in the regions of the latent space that they occupied. Further work will be required in order to make the method fully generalisable to light curves of other sources.

In regard to potential future work on the data set of GRS 1915+105 X-ray light curves, a supervised approach could be taken to study the link between the observed light curve patterns and underlying physics. Huppenkothen et al. (2017) used a classification scheme of GRS 1915+105 states into stochastic and chaotic processes (Harikrishnan et al., 2011) in order to leverage their machine learning classification algorithm in the study of the long term evolution of accretion properties of the source, and similar studies could be complemented using the method of feature extraction presented here. Ricketts et al. (2023) have made initial steps in this direction; they utilised a variant of the autoencoder-based feature extraction pipeline to analyse the same data set of GRS 1915+105 observations. They found a potential link between the variability classes δ and ϕ, as well as between ν and α, and recommended further investigation into the potential physical origin of those similarities.

Further work on this pipeline will involve the use of data augmentation to improve the ability of the network to generalise and interpolate between variability shapes, and hence increase the overlap of similar shapes from different sources in the latent space. The major differences between GRS 1915+105 and IGR J17091−3624 data lie in the frequency of signals, signal-to-noise ratio, and base intensity level. Additional light curves could be synthesised based on the GRS 1915+105 data, where the frequency of signals, noise levels and other parameters were modulated, and the LSTM-VAE could be trained on this augmented data set in order to expose it to a wide range of possible light curve shapes, and make the LSTM-VAE latent space more robust.

5.2 Long-term photometric variability

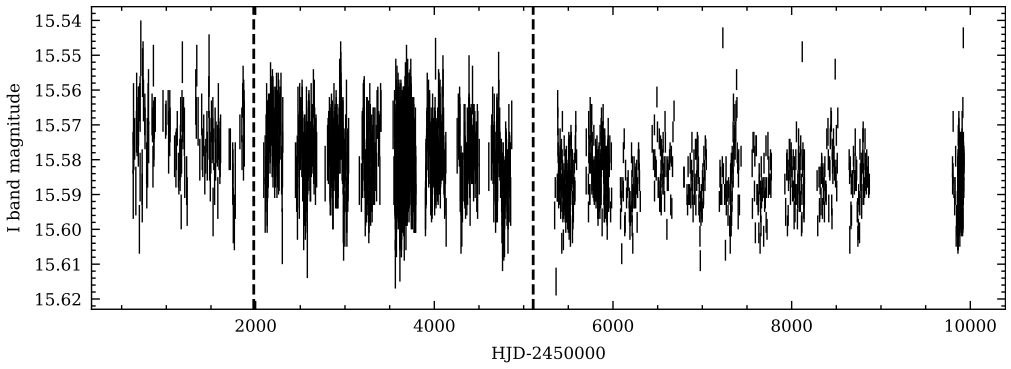
Chapter 4 introduces the wavelet-based pipeline of periodicity detection and characterisation, which can be a useful tool in the analysis of variability of Be X-ray binaries in the time scales of days, months, and years. After tuning the parameters of the method on synthetic data with a realistic observing pattern, I applied it to a sample of OGLE-XROM sources, hence detecting and characterising several new periodicities for the first time. These periodicities included the FRED-like modulations in three sources: SXP22.07, SXP304 and SXP726. In spite of the clearly asymmetric profile of the last detection, more photometric data is required in order to confirm that the signal is real. I also detected three new sinusoidal periodicities in the light curves of SXP9.13, SXP101 and SXP565. These periodicities are an indication of non-radial pulsations, and Lomb-Scargle analysis of these detections revealed the likely period of NRPs in those sources.

The method has potential to be applied to larger samples of Be X-ray binaries, which could help to quantify the correlation between non-radial pulsations and the swelling of the Be decretion disc, in order to shed some light on the potential causality between these phenomena.

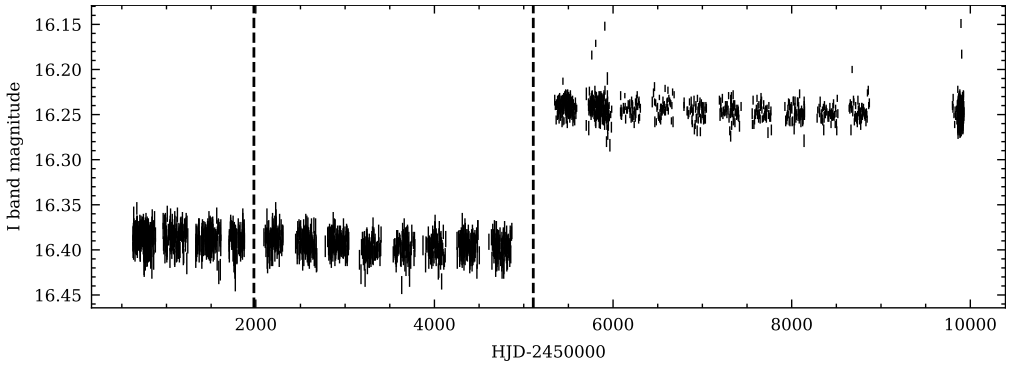
Appendix A

Light curves and WWZ scalograms of OGLE-XROM photometry

A.1 Light curves

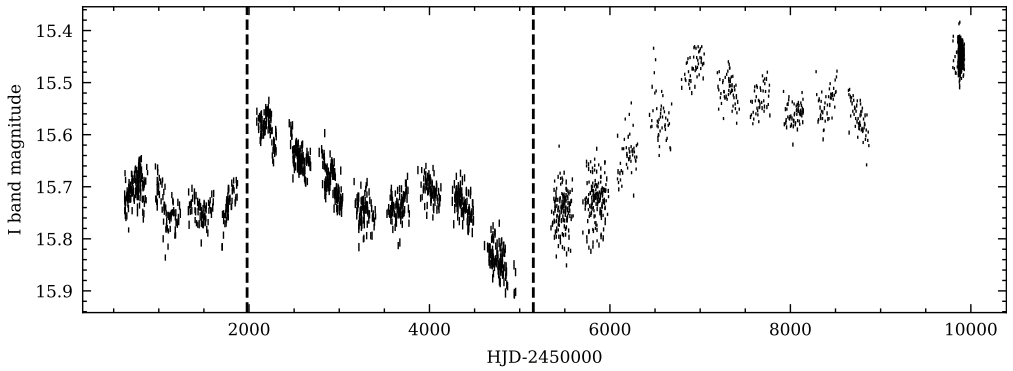


(a) SXP0.09

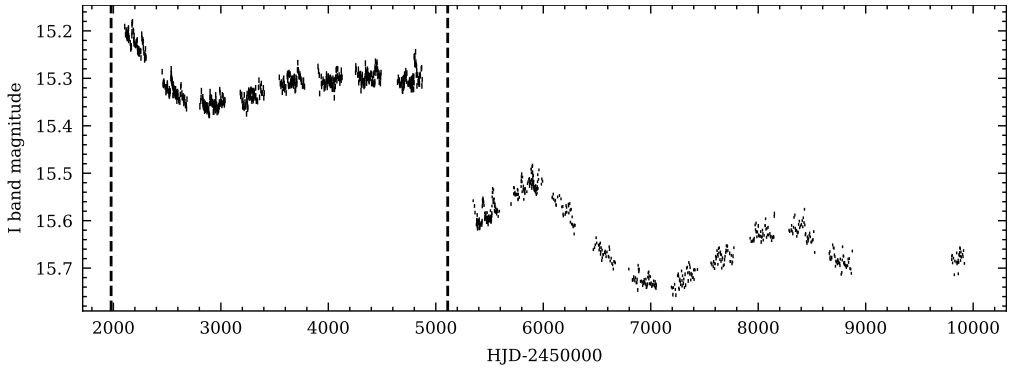


(b) SXP0.92

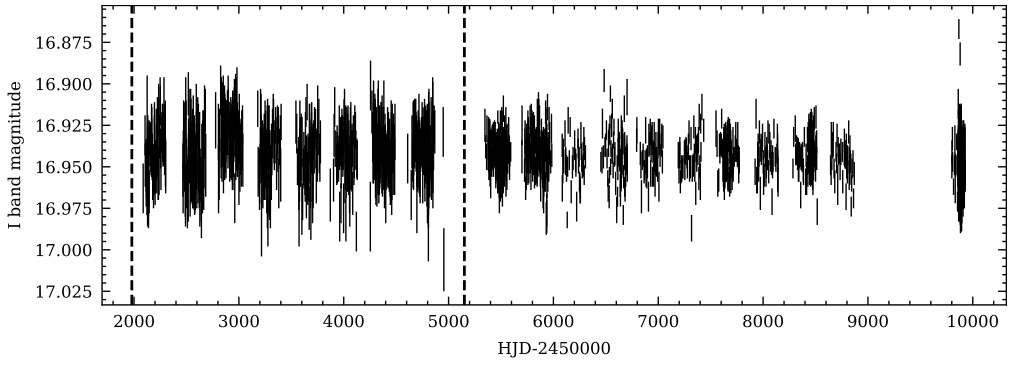
FIGURE A.1: I-band OGLE light curve of analysed sources. OGLE-II, III and IV are separated by black dashed lines.



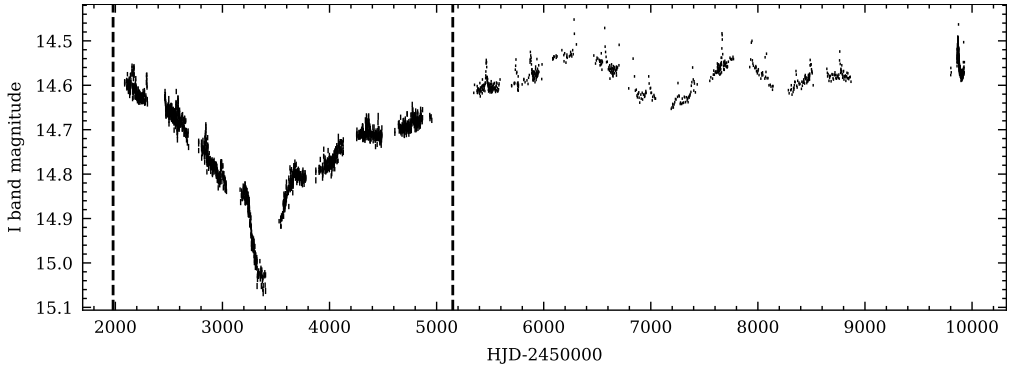
(a) SXP25.5



(b) SXP31.0

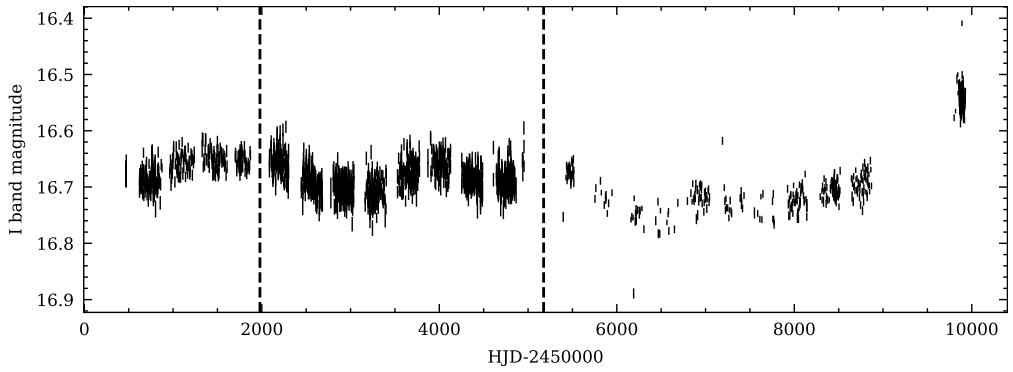


(c) SXP34.08

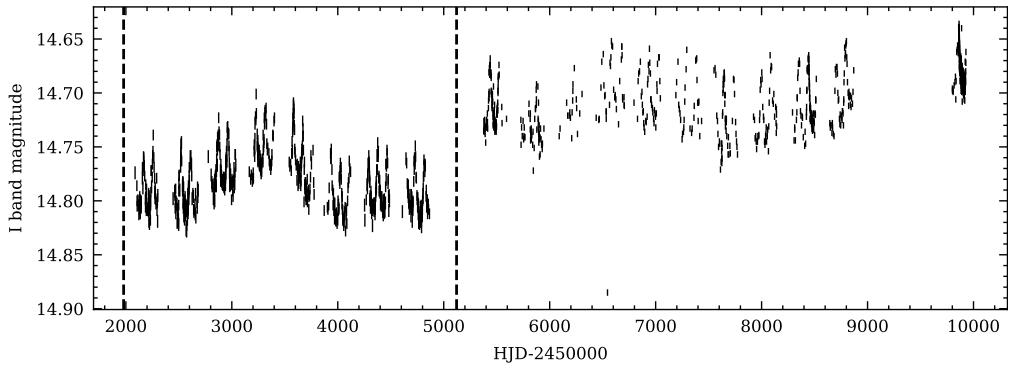


(d) SXP46.6

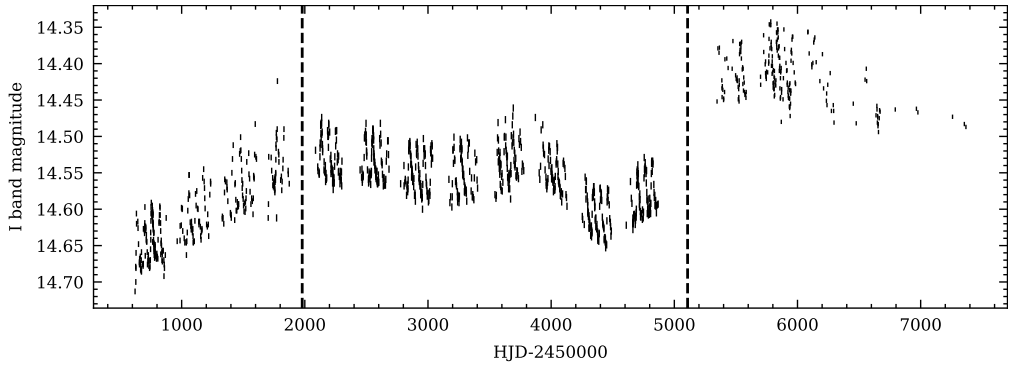
FIGURE A.2: I-band OGLE light curve of analysed sources. OGLE-II, III and IV are separated by black dashed lines.



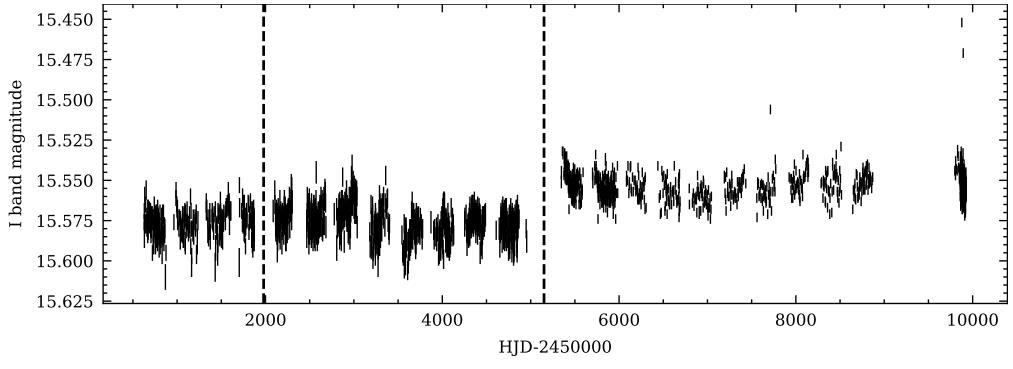
(a) SXP74.7



(b) SXP91.1

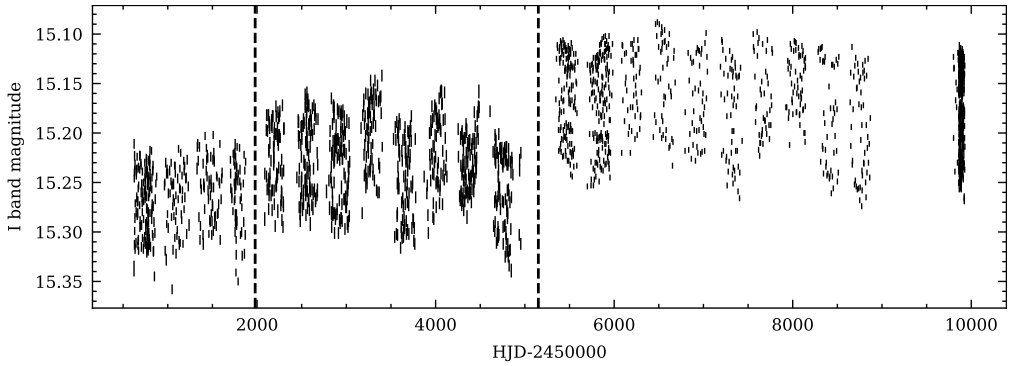


(c) SXP293

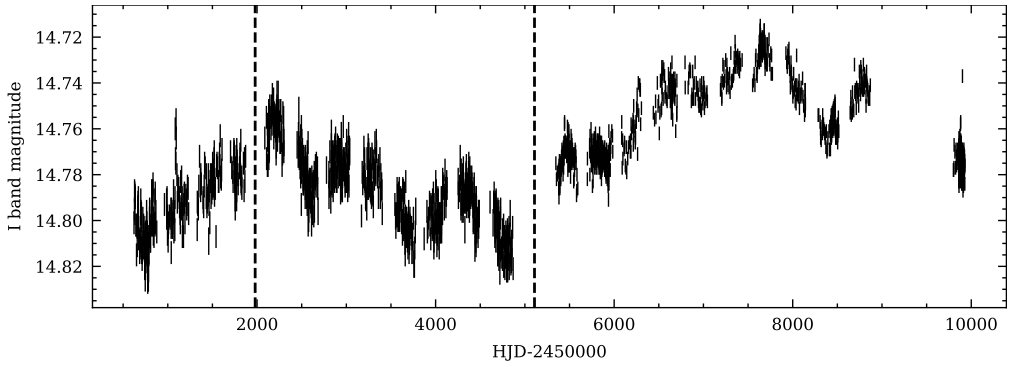


(d) SXP304

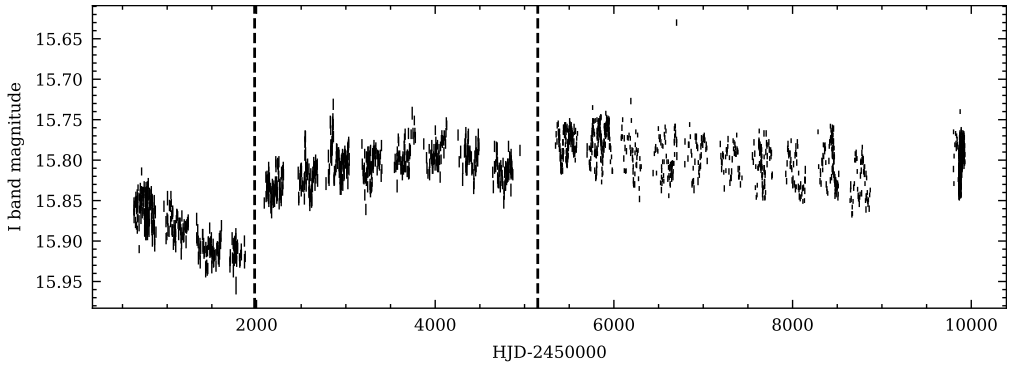
FIGURE A.3: I-band OGLE light curve of analysed sources. OGLE-II, III and IV are separated by black dashed lines.



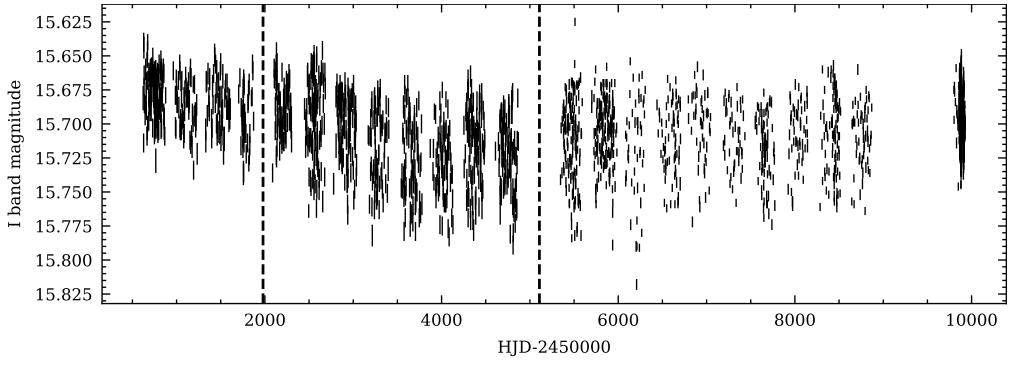
(a) SXP323



(b) SXP504

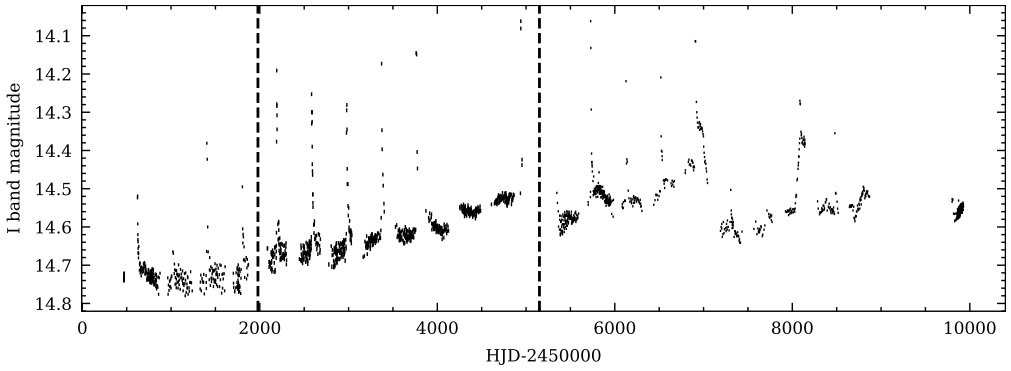


(c) SXP565

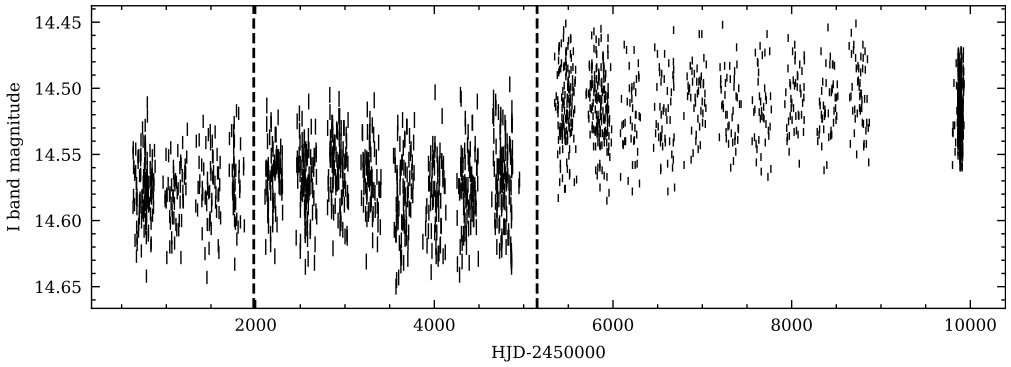


(d) SXP701

FIGURE A.4: I-band OGLE light curve of analysed sources. OGLE-II, III and IV are separated by black dashed lines.



(a) SXP756



(b) SXP1323

FIGURE A.5: I-band OGLE light curve of analysed sources. OGLE-II, III and IV are separated by black dashed lines.

A.2 WWZ transform of synthetic data

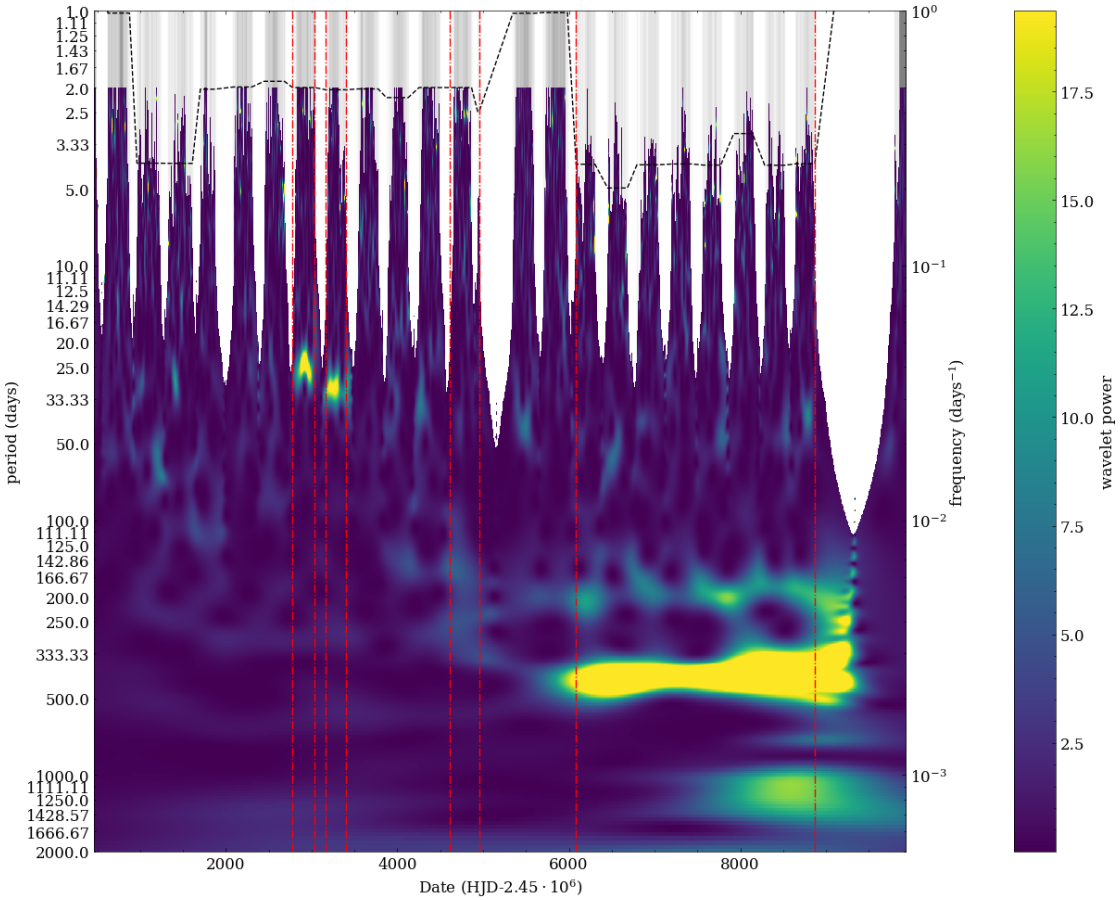


FIGURE A.6: An alternative version of the WWZ transform from Figure 4.1b, with decay constant $c = 1/(8\pi^2)$ and not N_{eff} filtering. It is a WWZ transform of a synthetic light curve with the typical OGLE sampling, and transient signals of 25, 30, 2.5 and 400 days injected into its separate subsections (see Figure 4.1a). Red dash-dotted boxes indicate signals as above. Grey vertical lines indicate the observation timestamps. The series of horizontal dashed black lines shows the quasi-Nyquist frequency, calculated as the median time difference between time stamps in every observing season. The wavelet power value is capped at 19.5 (corresponding to $\alpha = 0.05$ for an F-distribution with ∞ and 2 degrees of freedom).

A.3 WWZ transforms of OGLE-XROM light curves

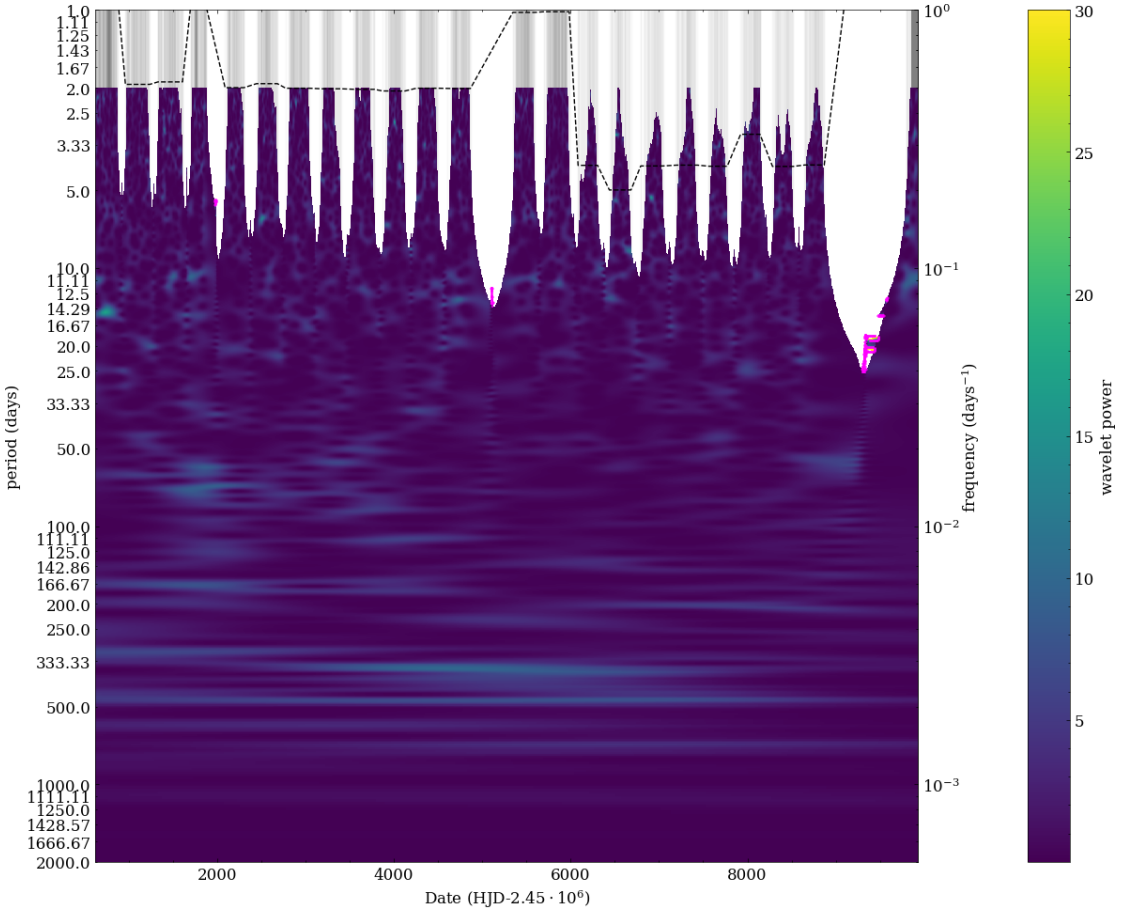


FIGURE A.8: WWZ transform of SXP0.92 light curve.

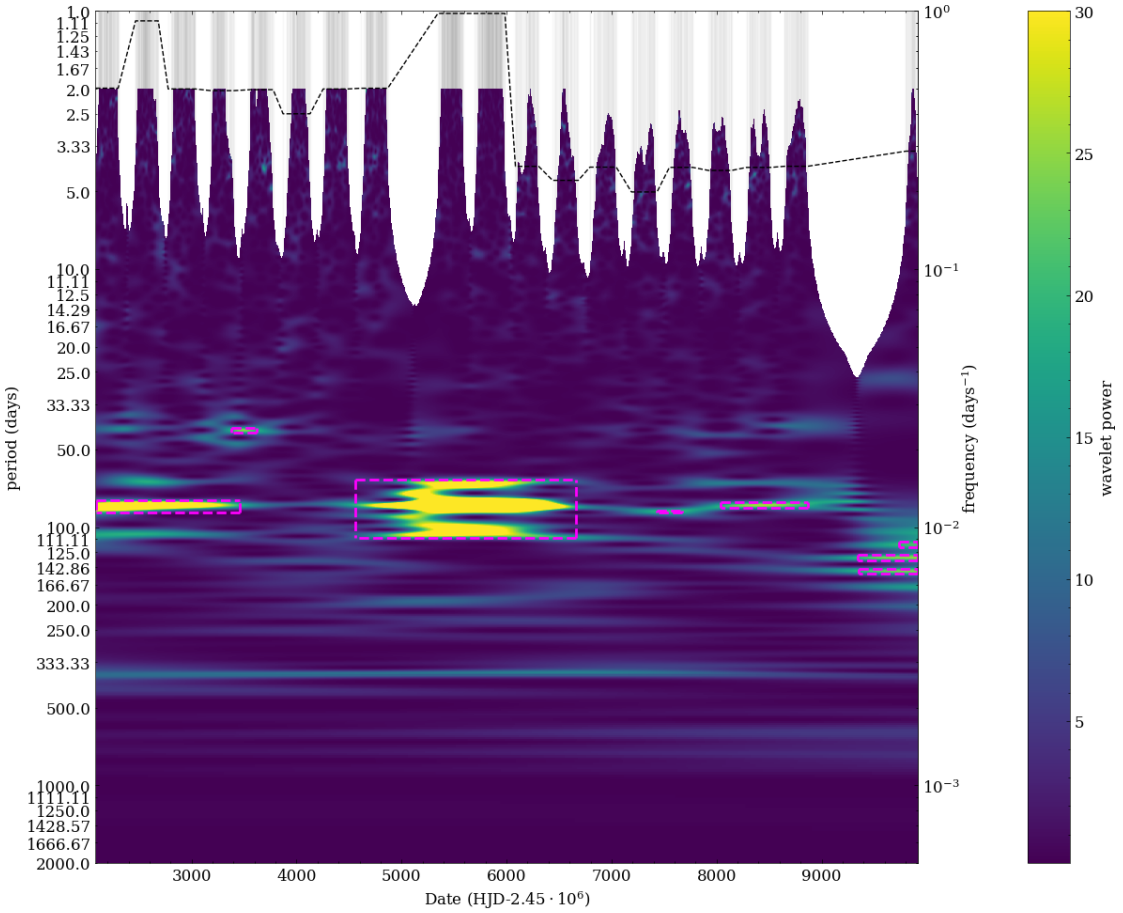


FIGURE A.9: WWZ transform of SXP2.763 light curve.

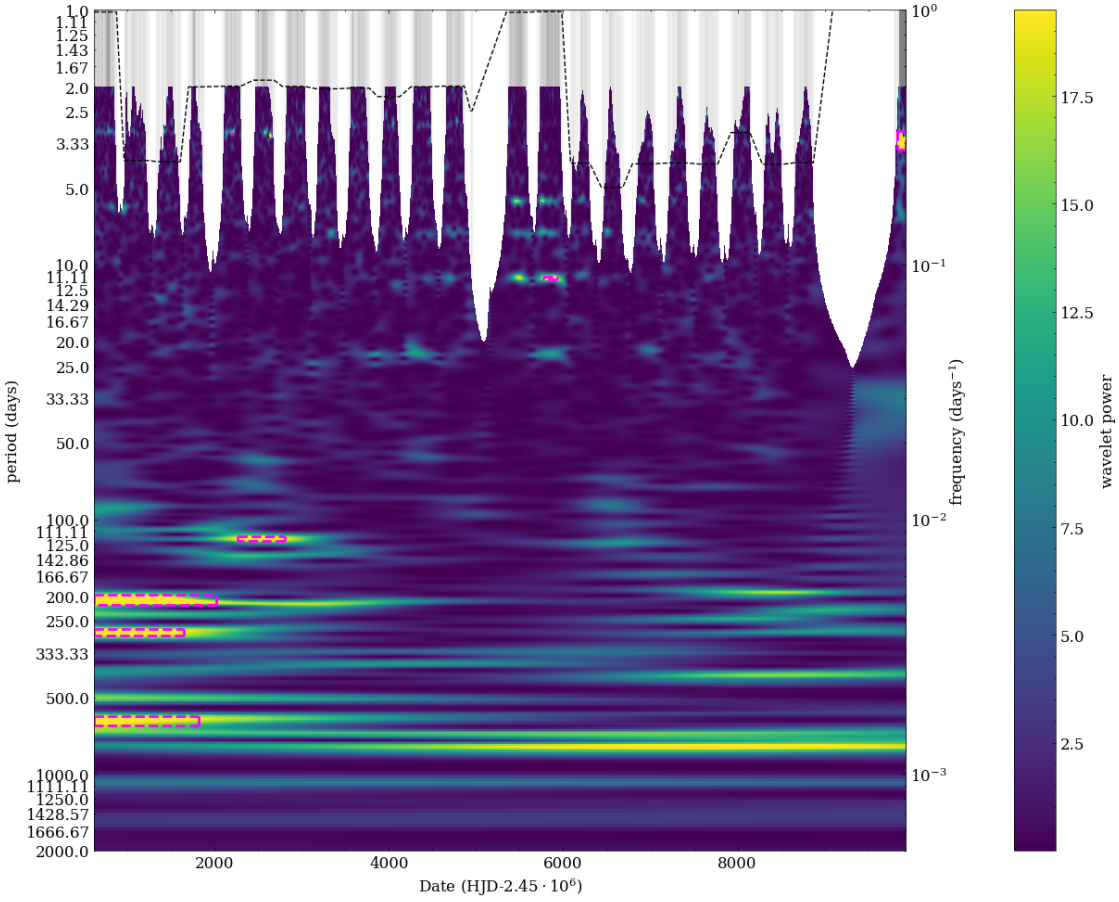


FIGURE A.10: WWZ transform of SXP25.5 light curve.

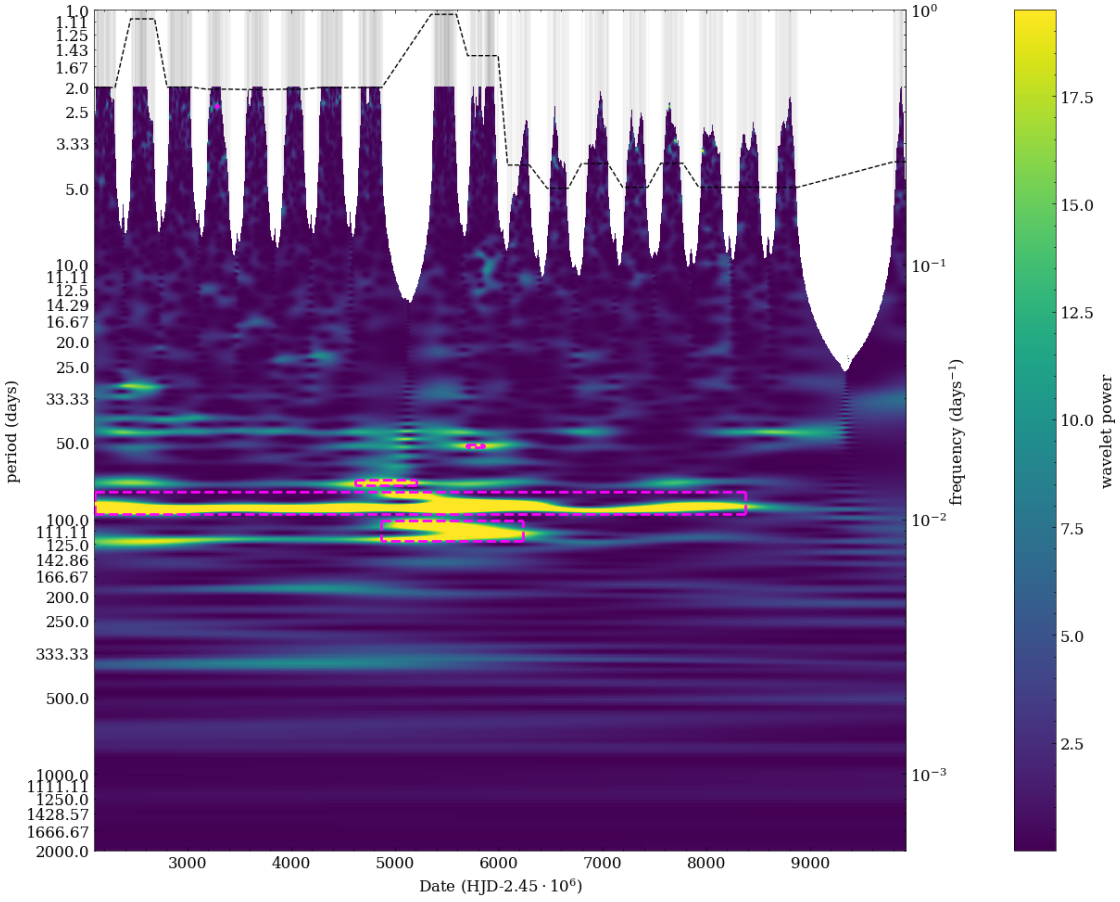


FIGURE A.11: WWZ transform of SXP31.0 light curve.

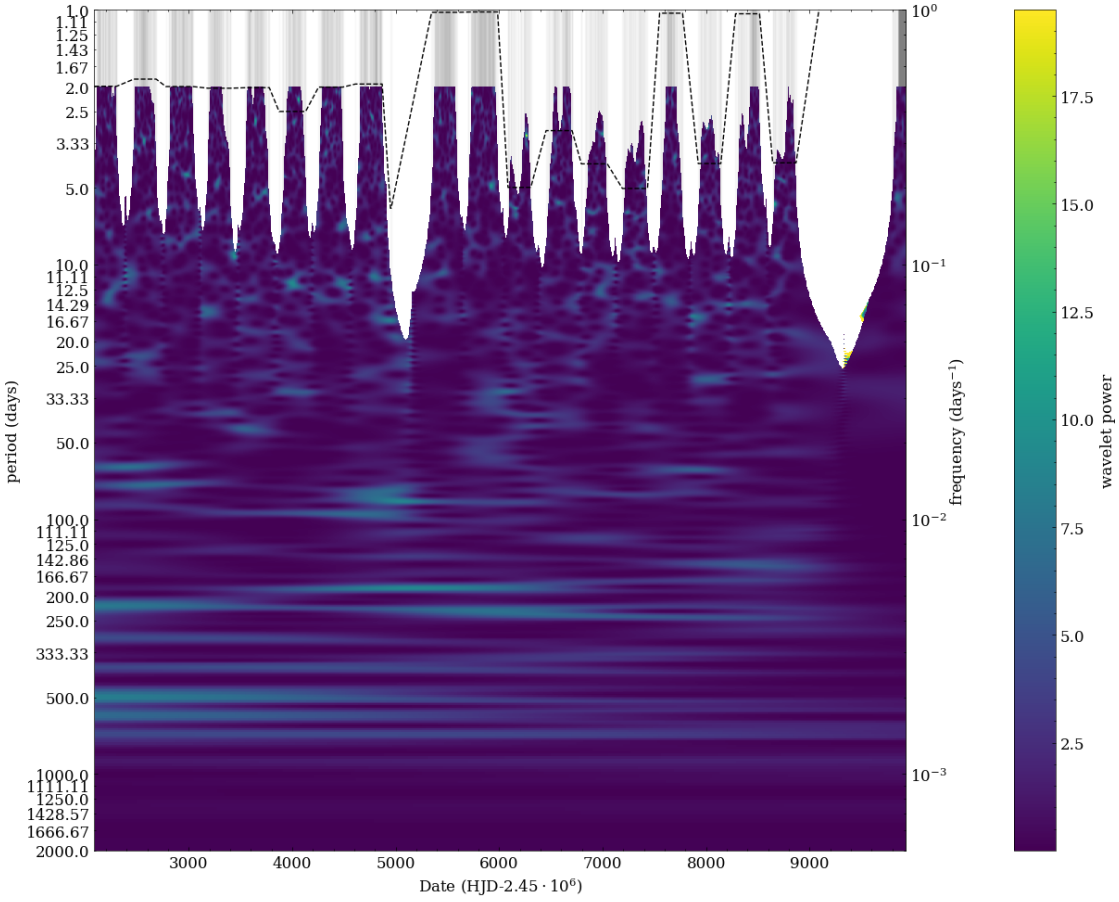


FIGURE A.12: WWZ transform of SXP34.08 light curve.

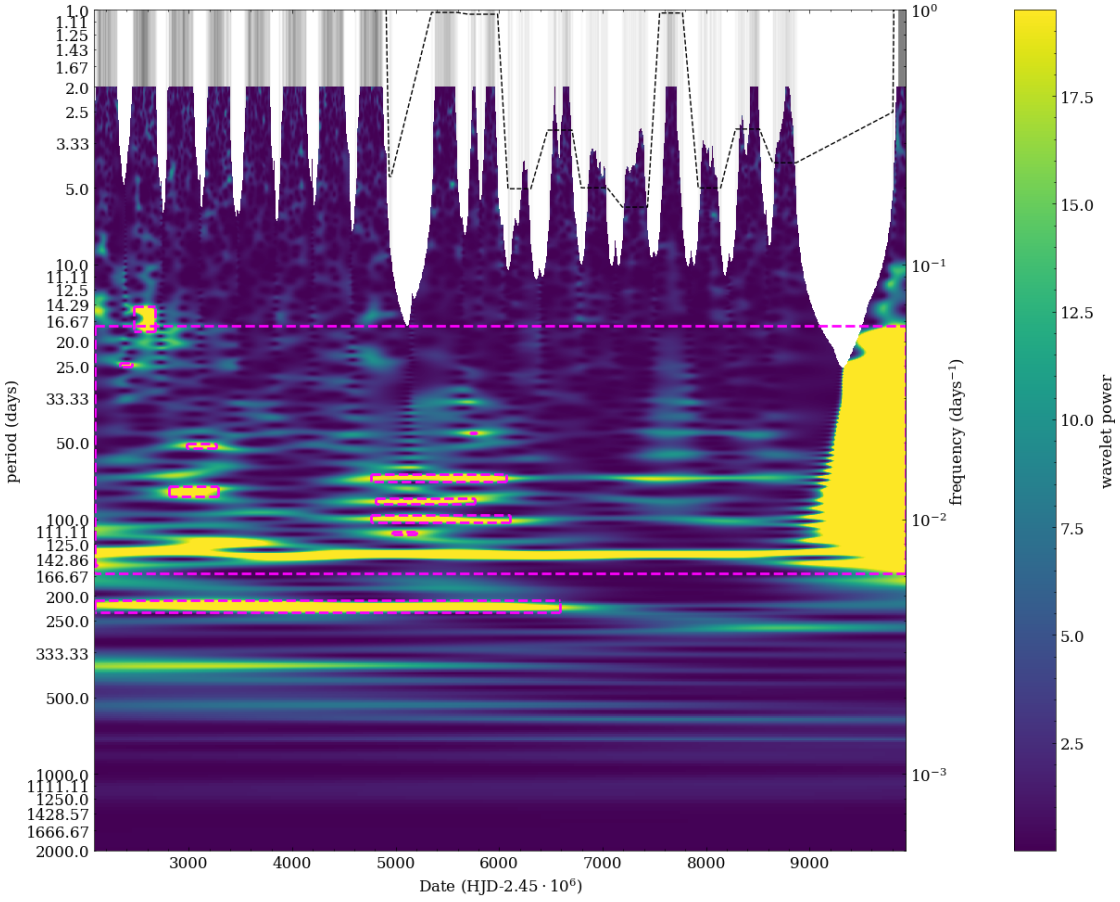


FIGURE A.13: WWZ transform of SXP46.6 light curve.

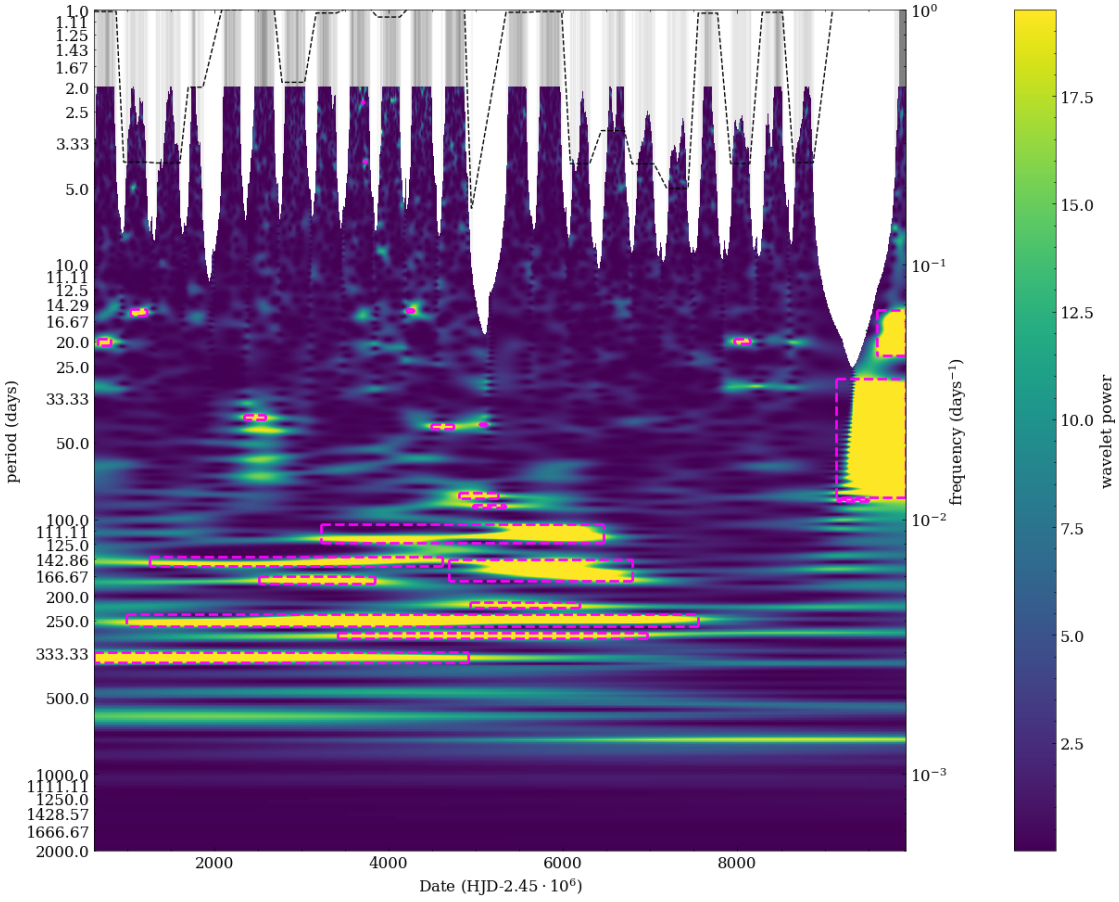


FIGURE A.14: WWZ transform of SXP59.0 light curve.

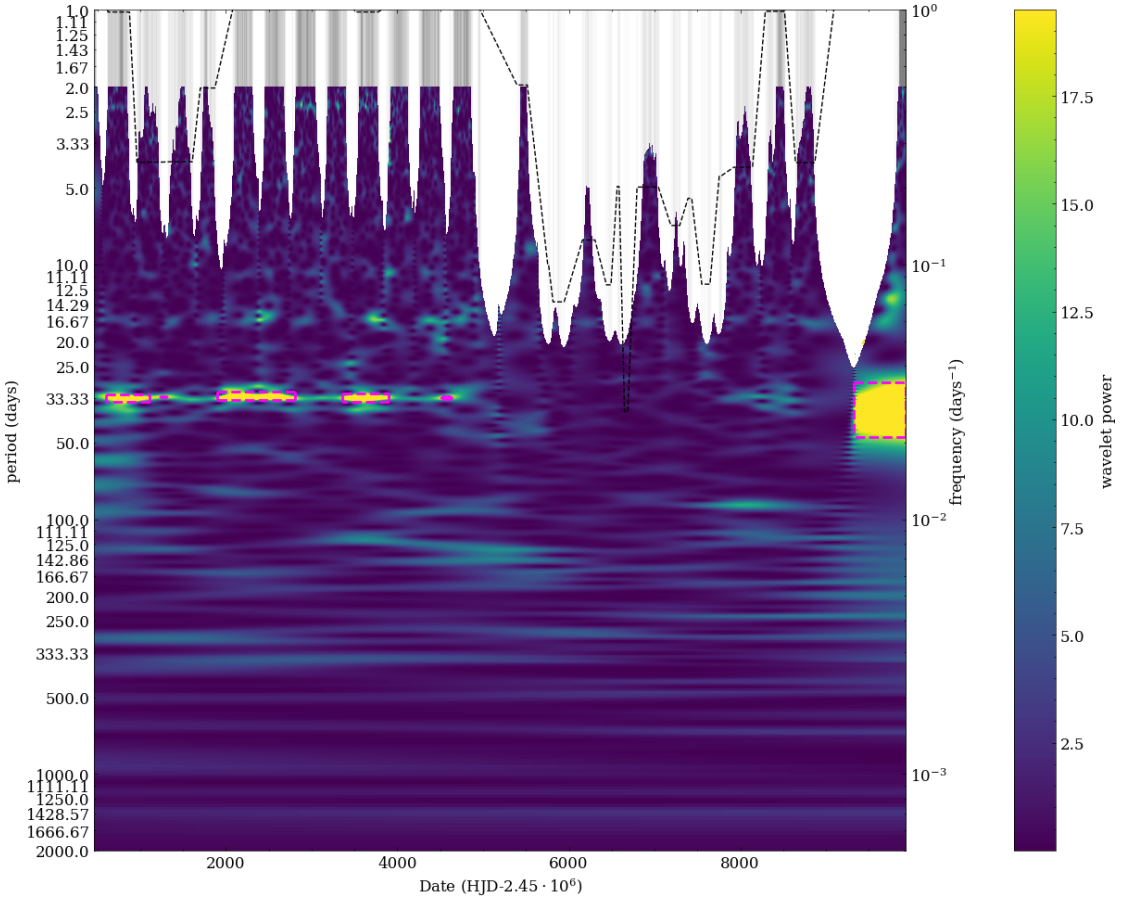


FIGURE A.15: WWZ transform of SXP74.7 light curve.

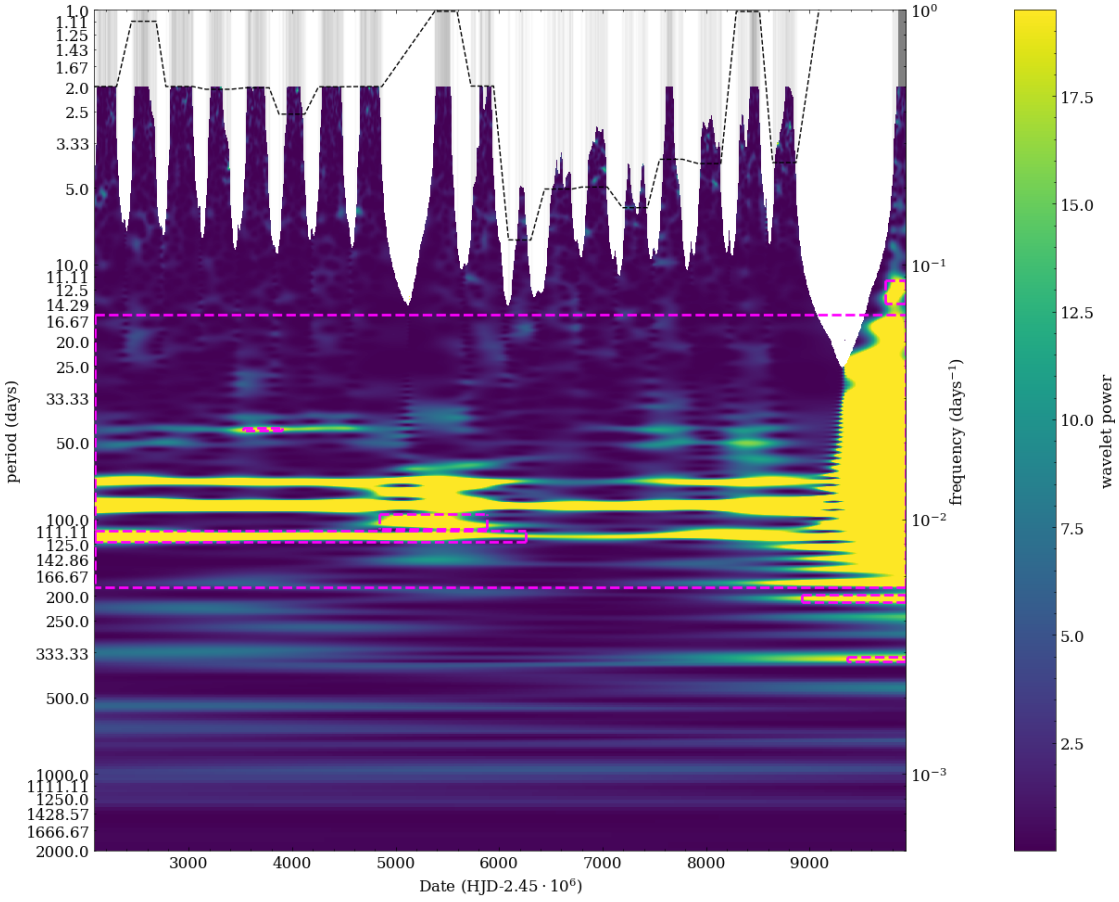


FIGURE A.16: WWZ transform of SXP91.1 light curve.

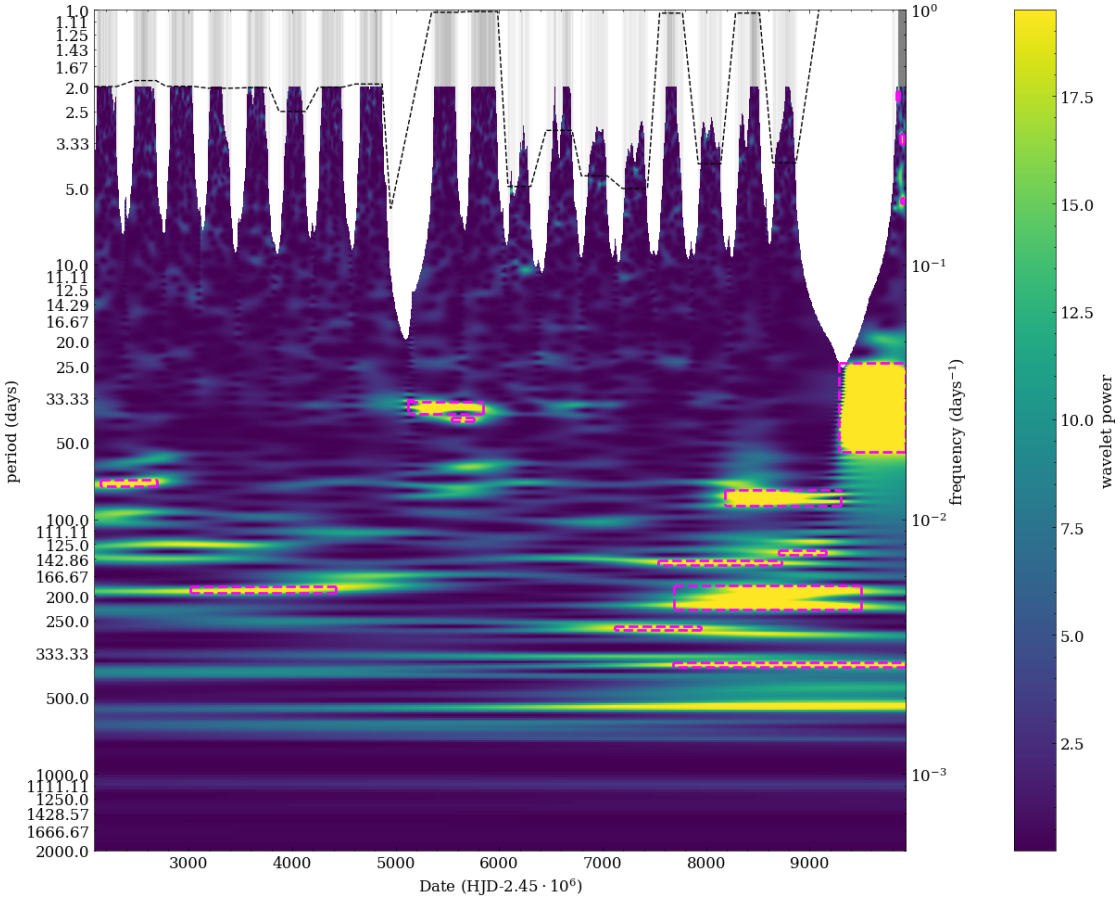


FIGURE A.17: WWZ transform of SXP152.1 light curve.

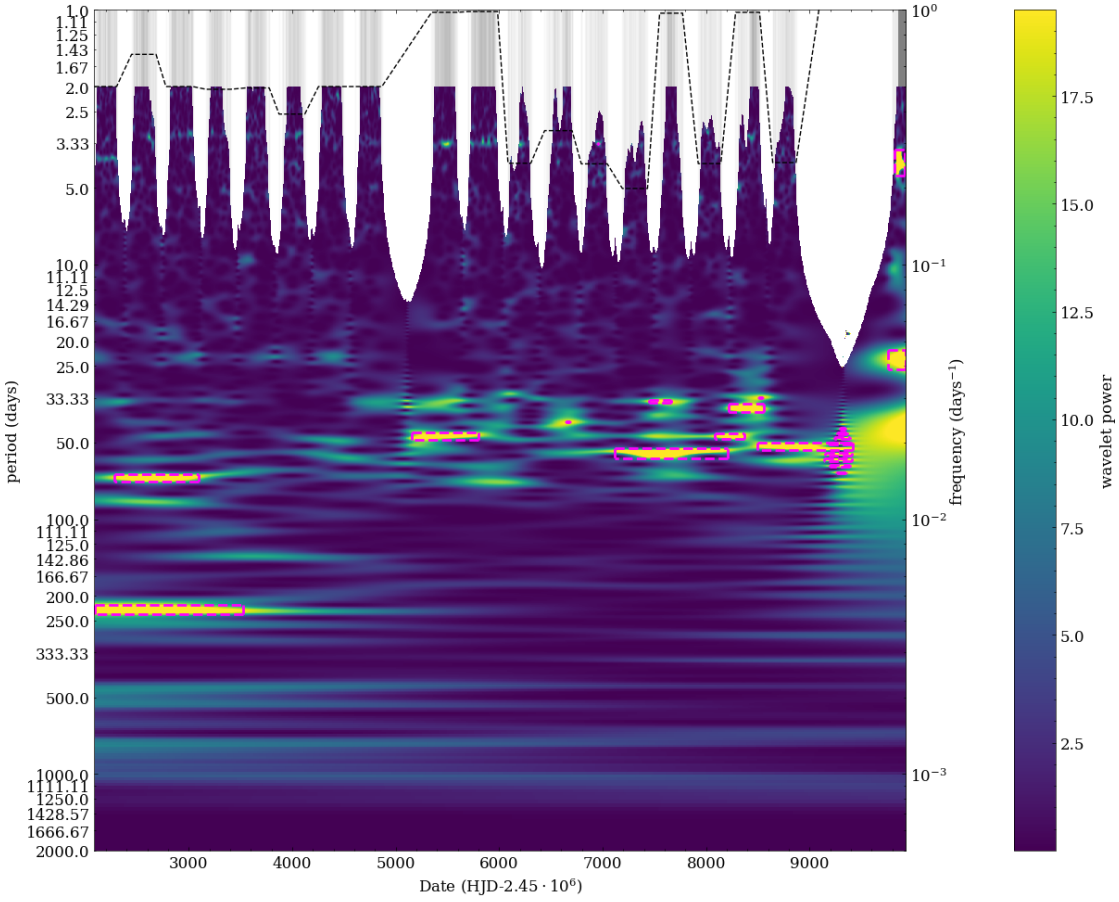


FIGURE A.18: WWZ transform of SXP169.3 light curve.

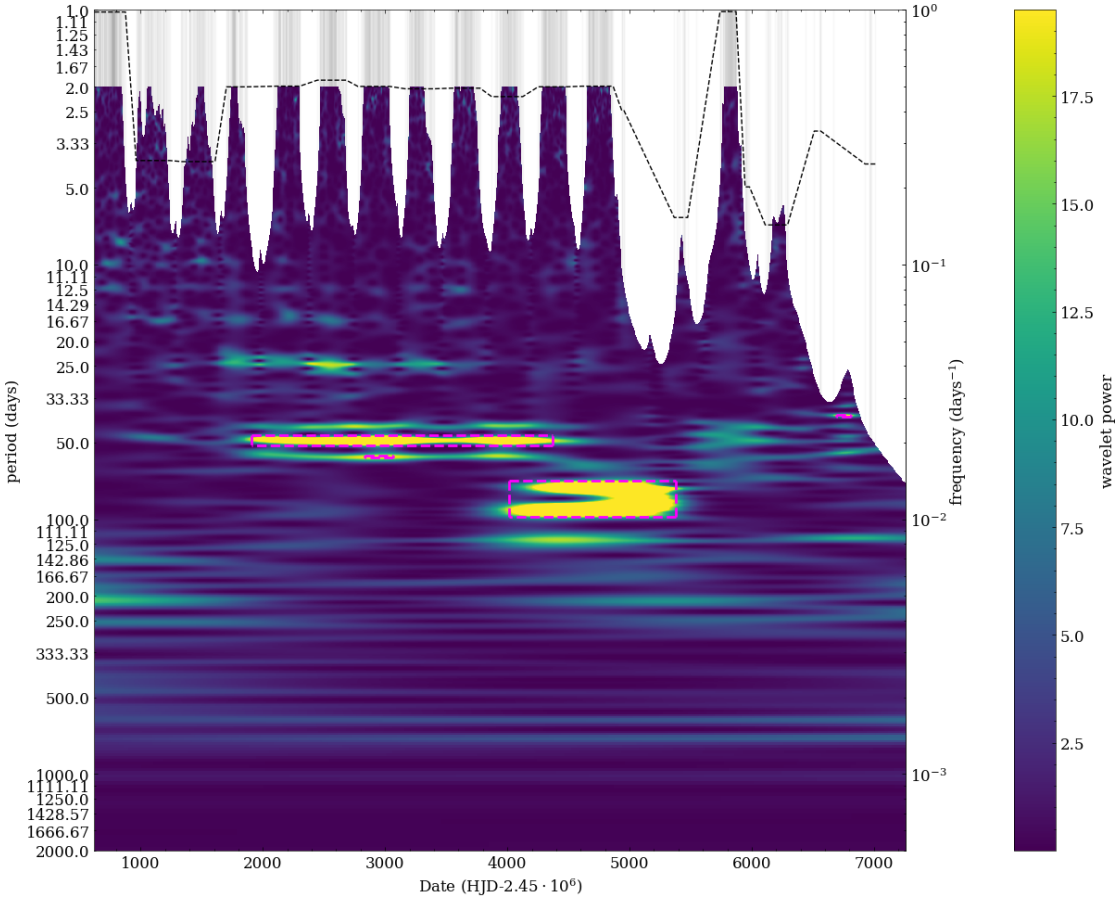


FIGURE A.19: WWZ transform of SXP264 light curve.

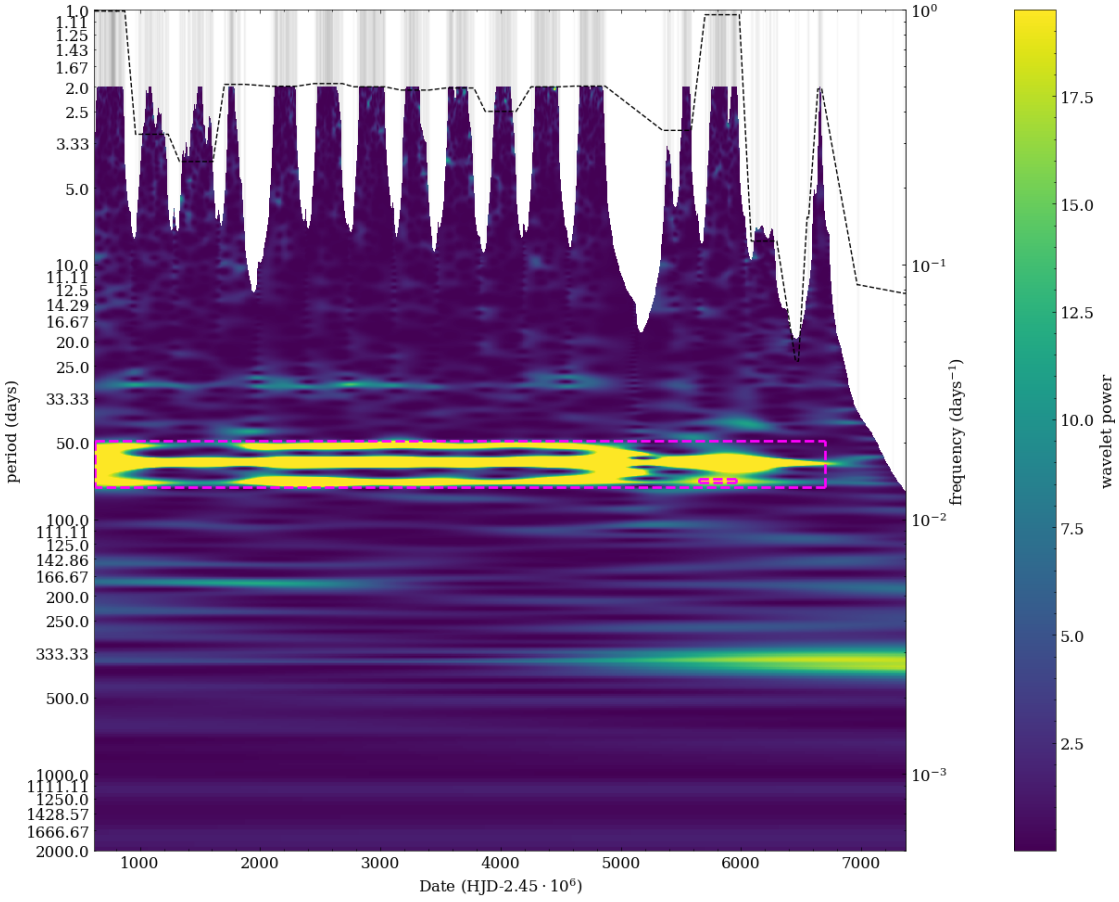


FIGURE A.20: WWZ transform of SXP293 light curve.

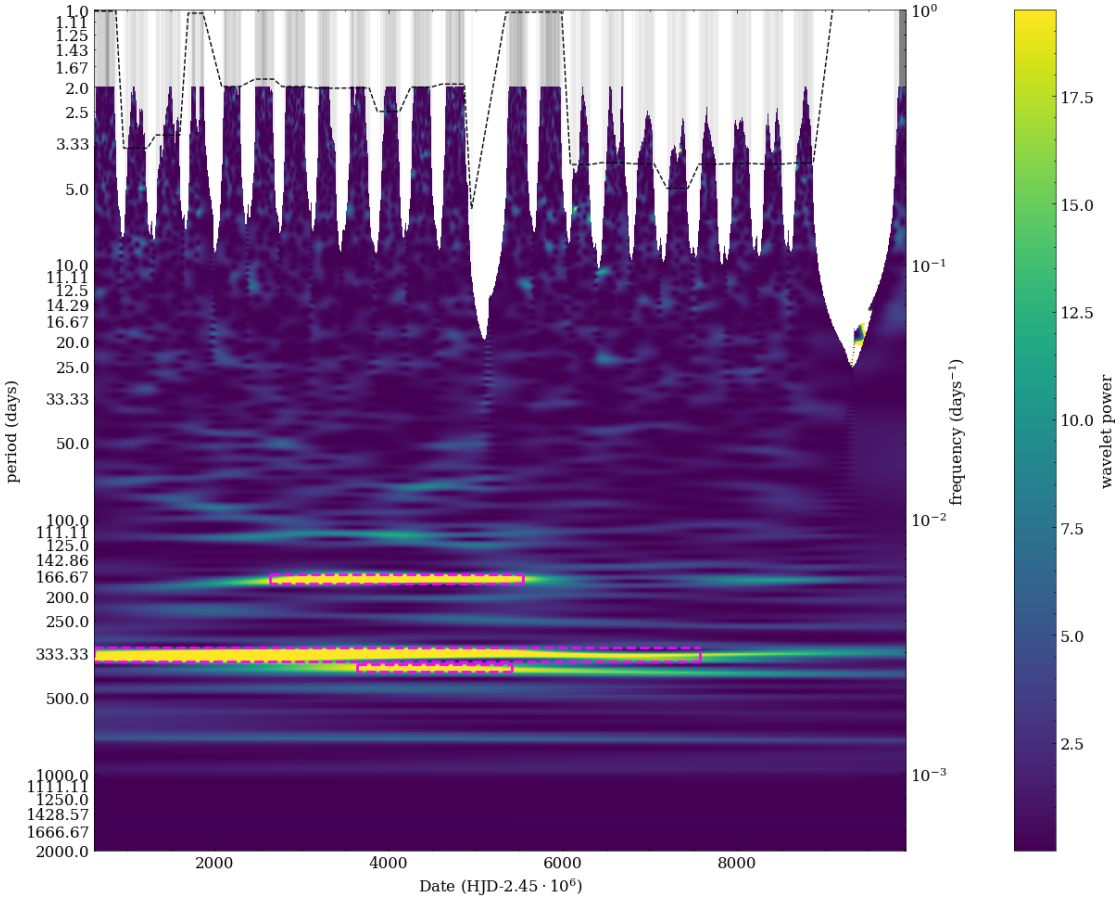


FIGURE A.21: WWZ transform of SXP304 light curve.

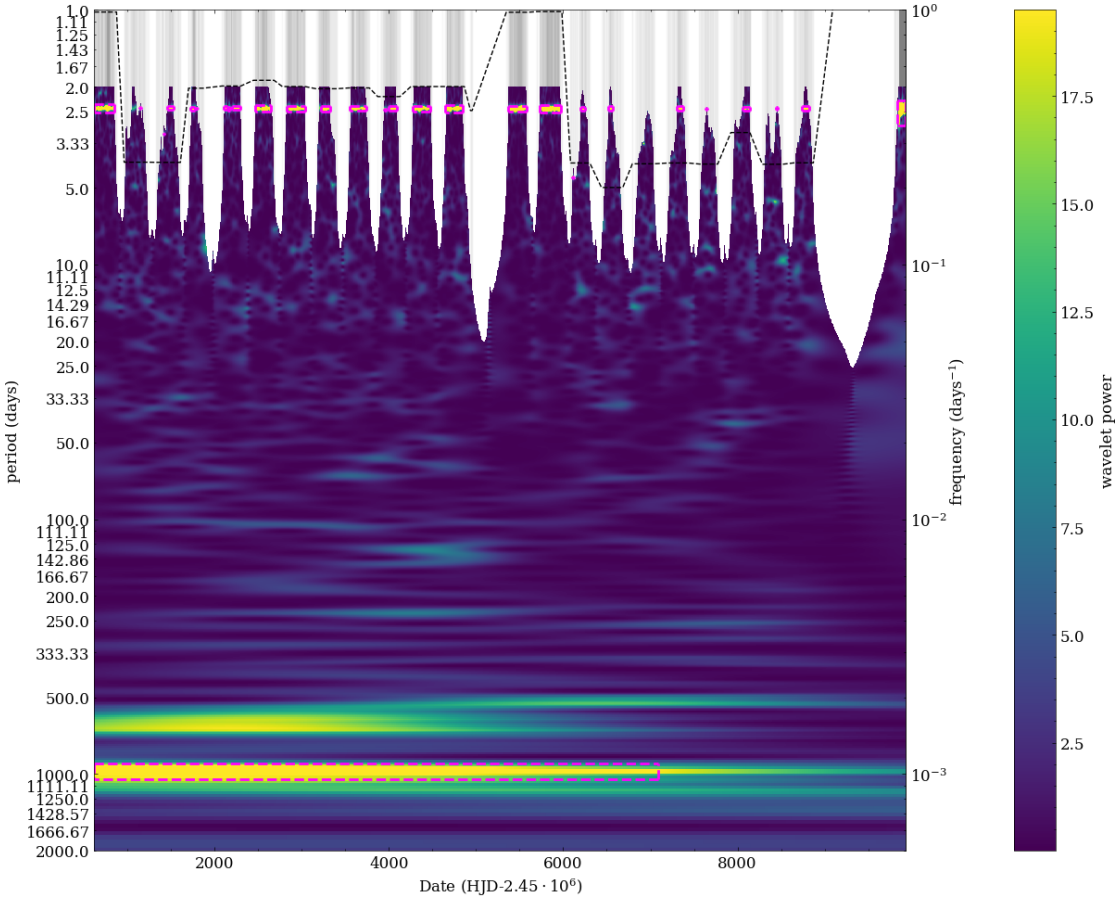


FIGURE A.22: WWZ transform of SXP323 light curve.

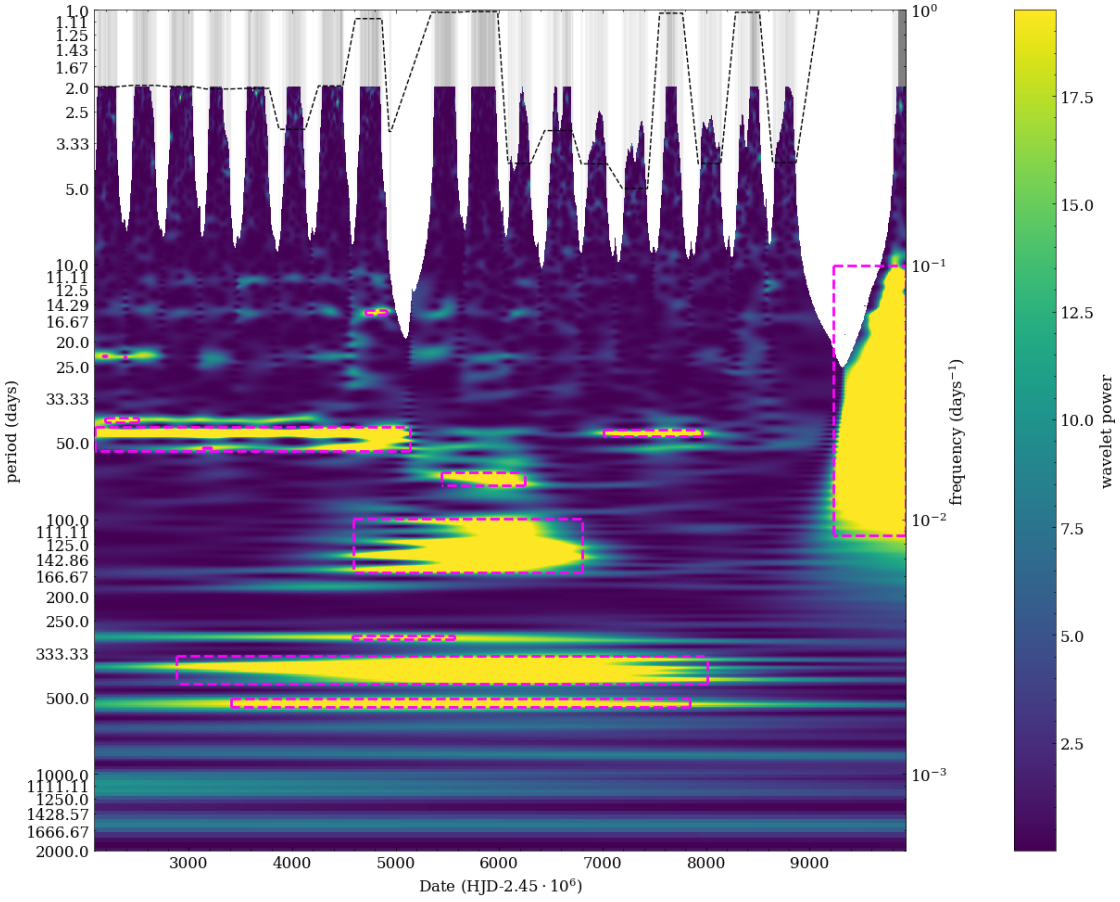


FIGURE A.23: WWZ transform of SXP327 light curve.

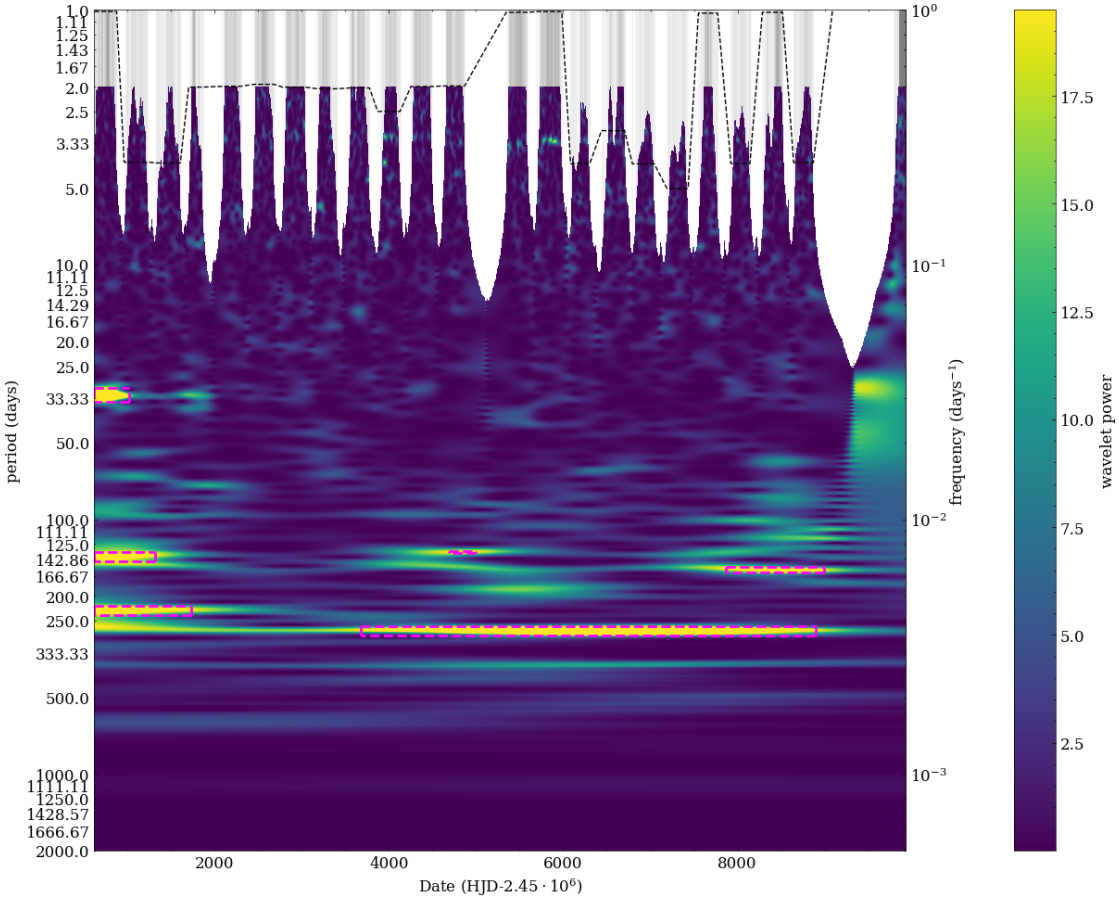


FIGURE A.24: WWZ transform of SXP504 light curve.

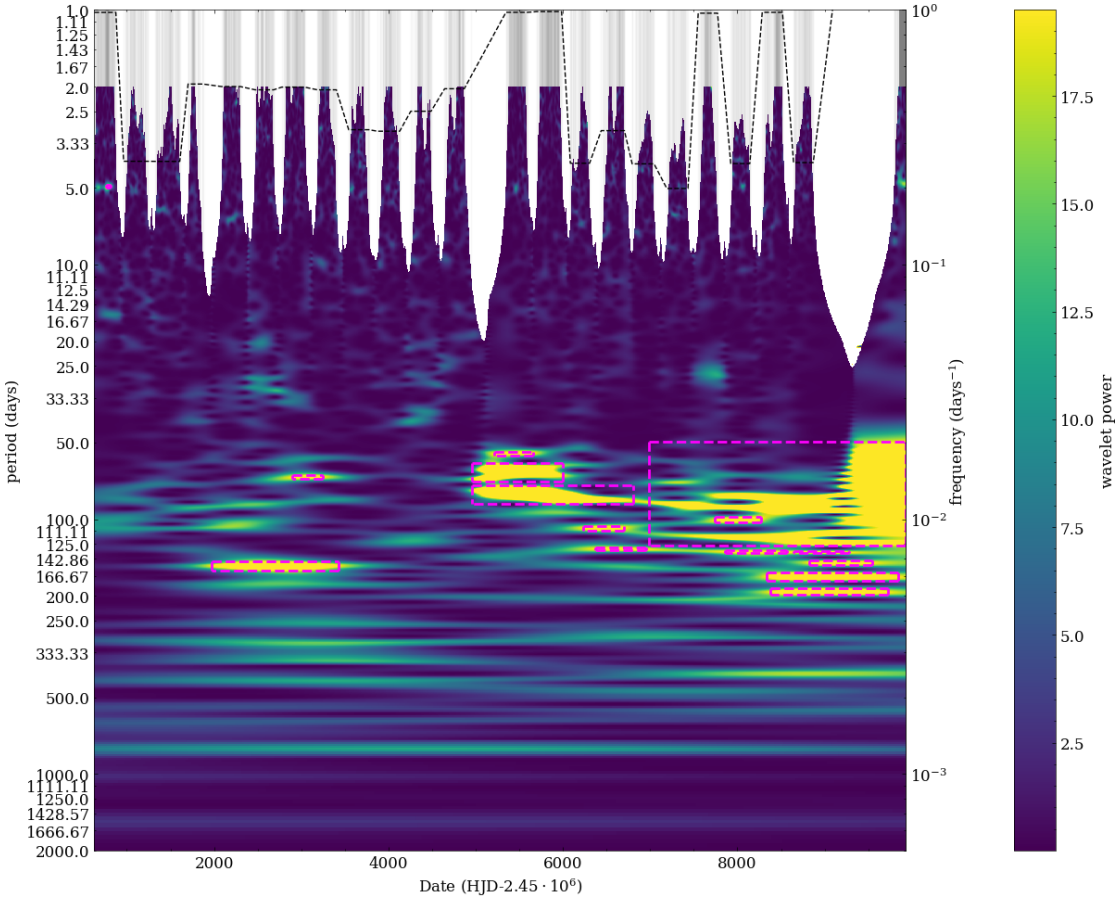


FIGURE A.25: WWZ transform of SXP565 light curve.

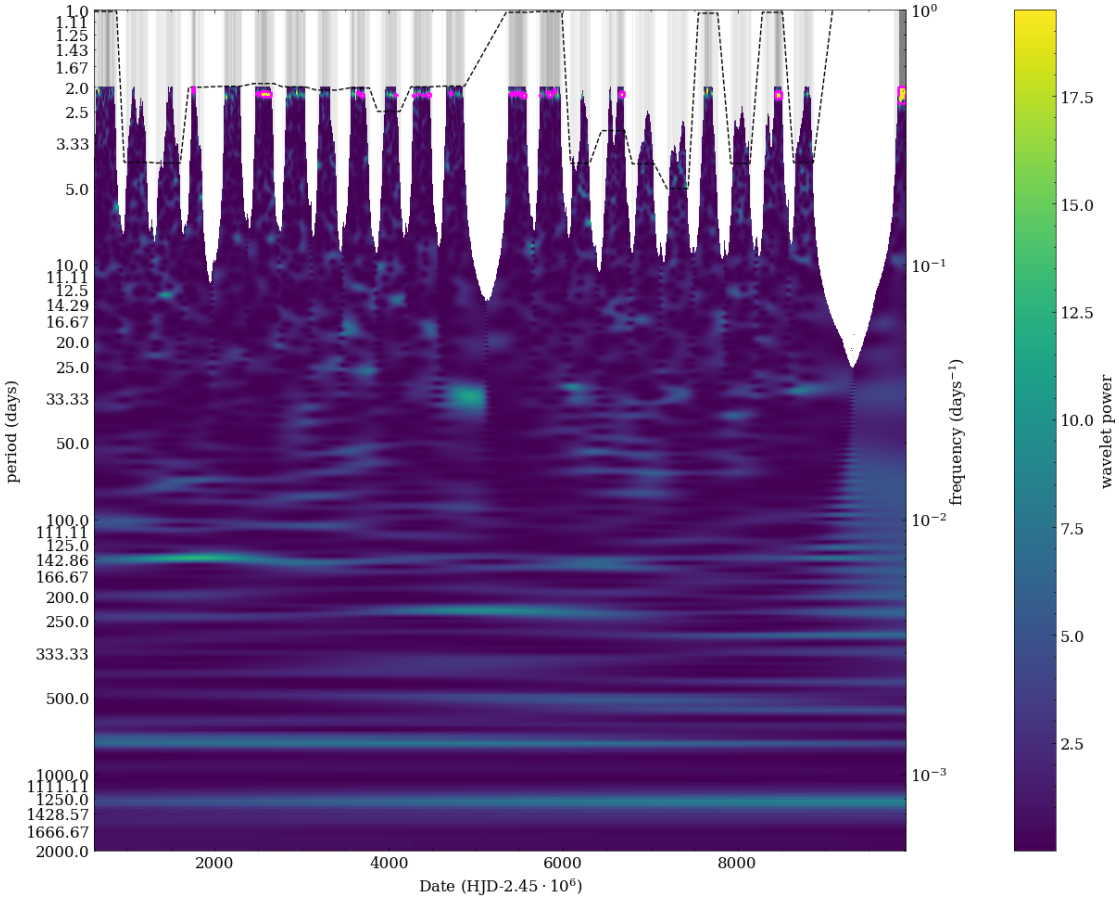


FIGURE A.26: WWZ transform of SXP701 light curve.

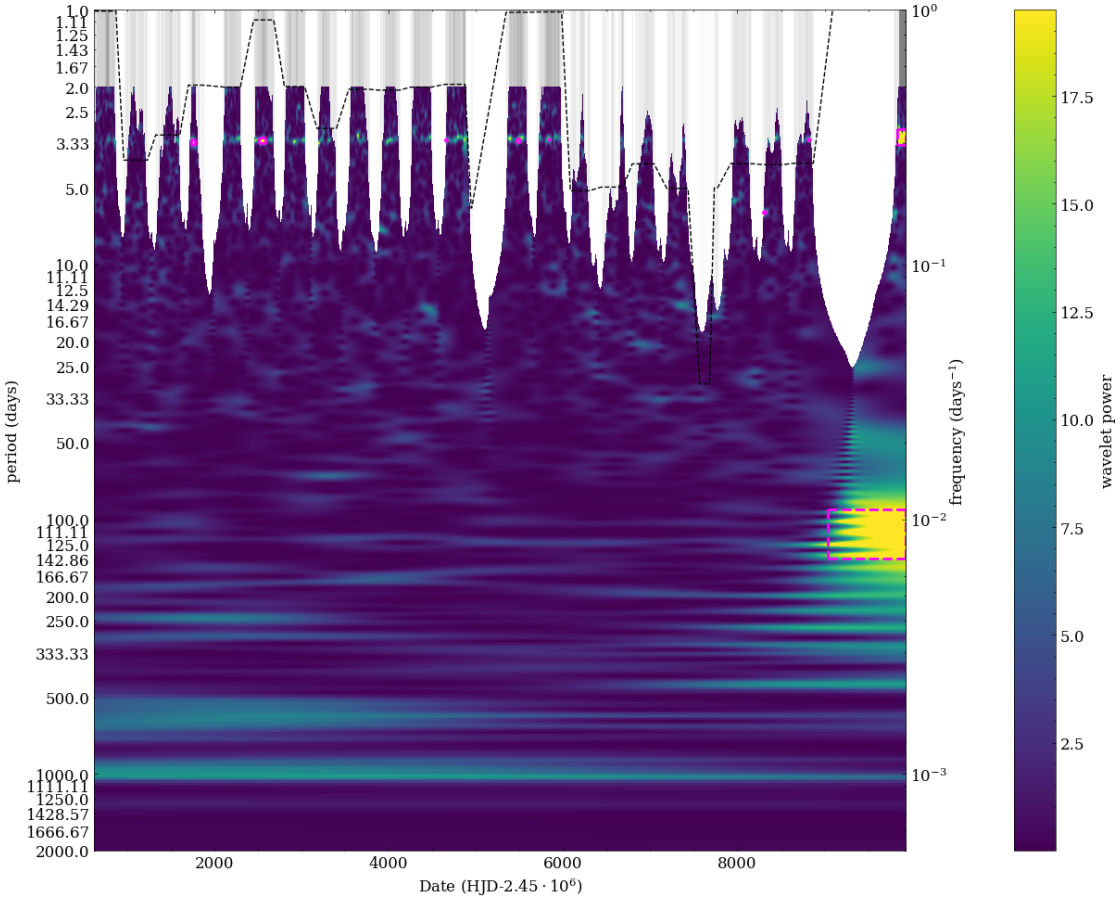


FIGURE A.27: WWZ transform of SXP726 light curve.

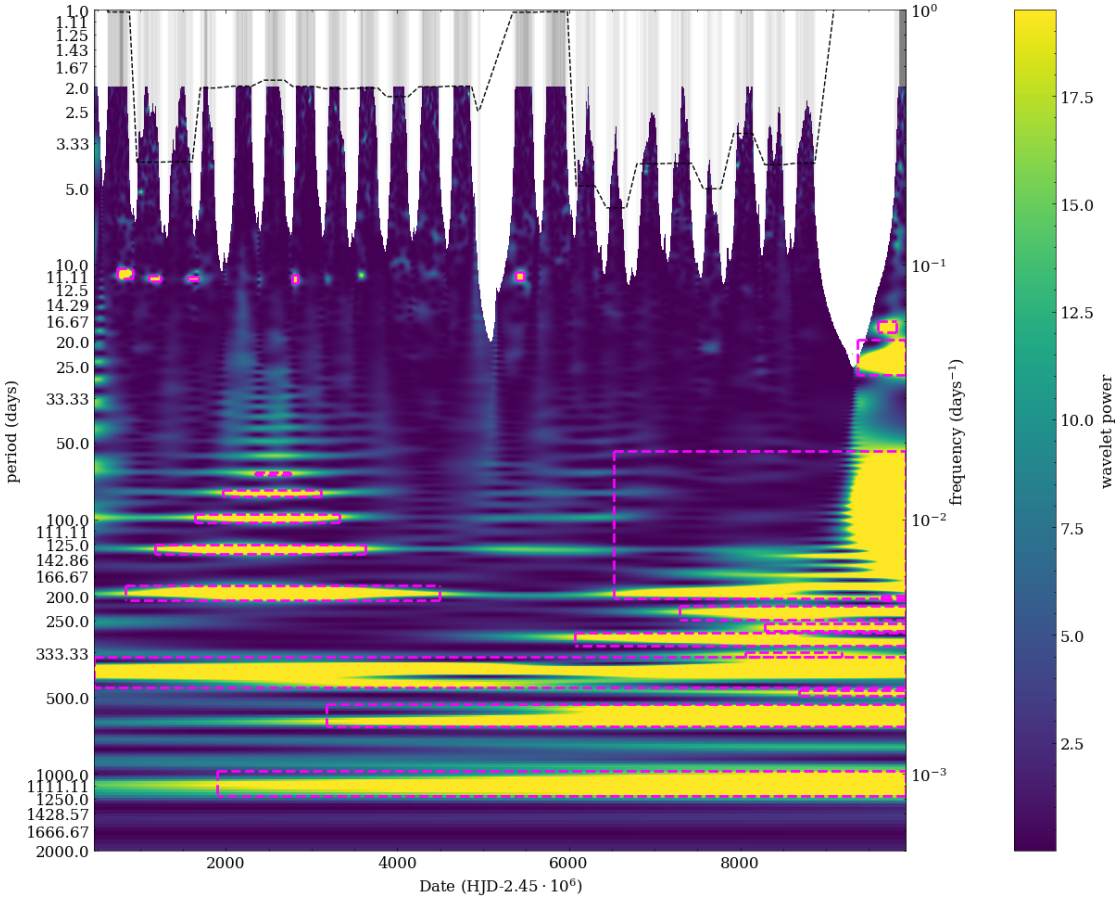


FIGURE A.28: WWZ transform of SXP756 light curve.

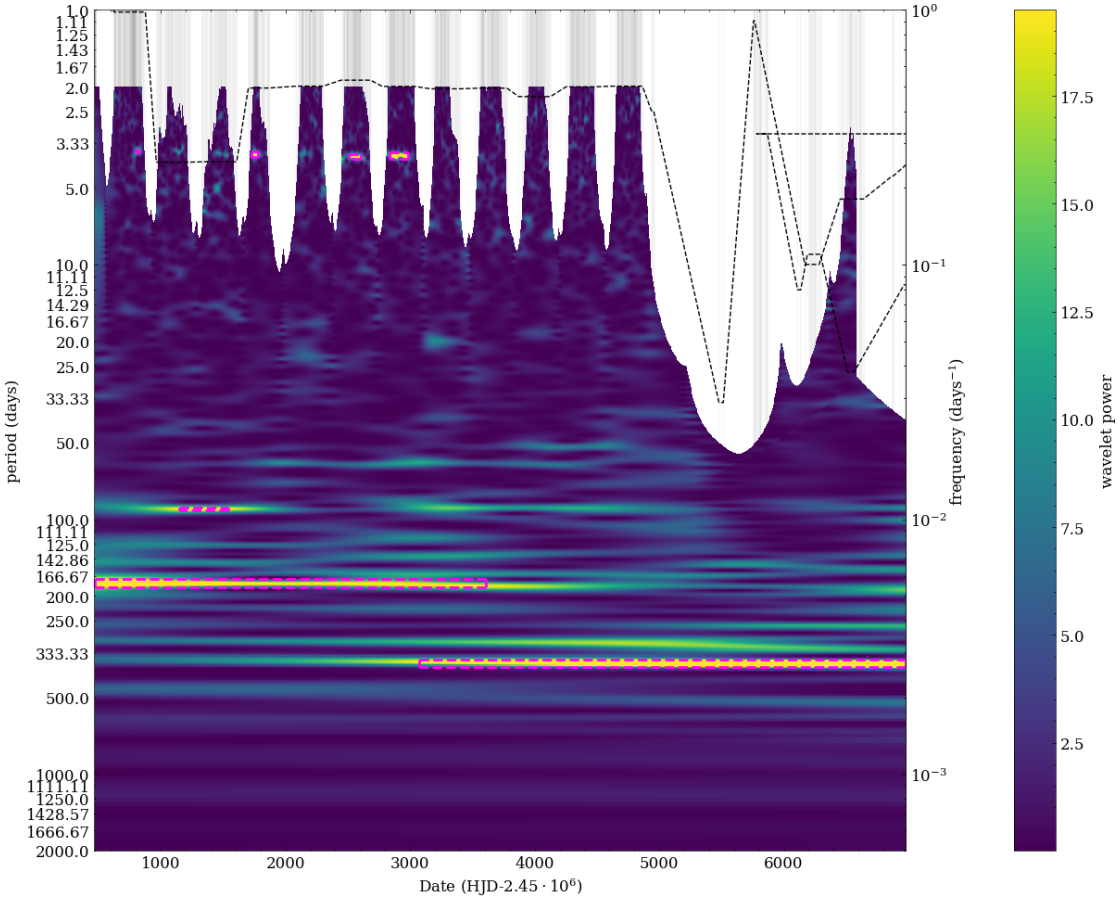


FIGURE A.29: WWZ transform of SXP893 light curve.

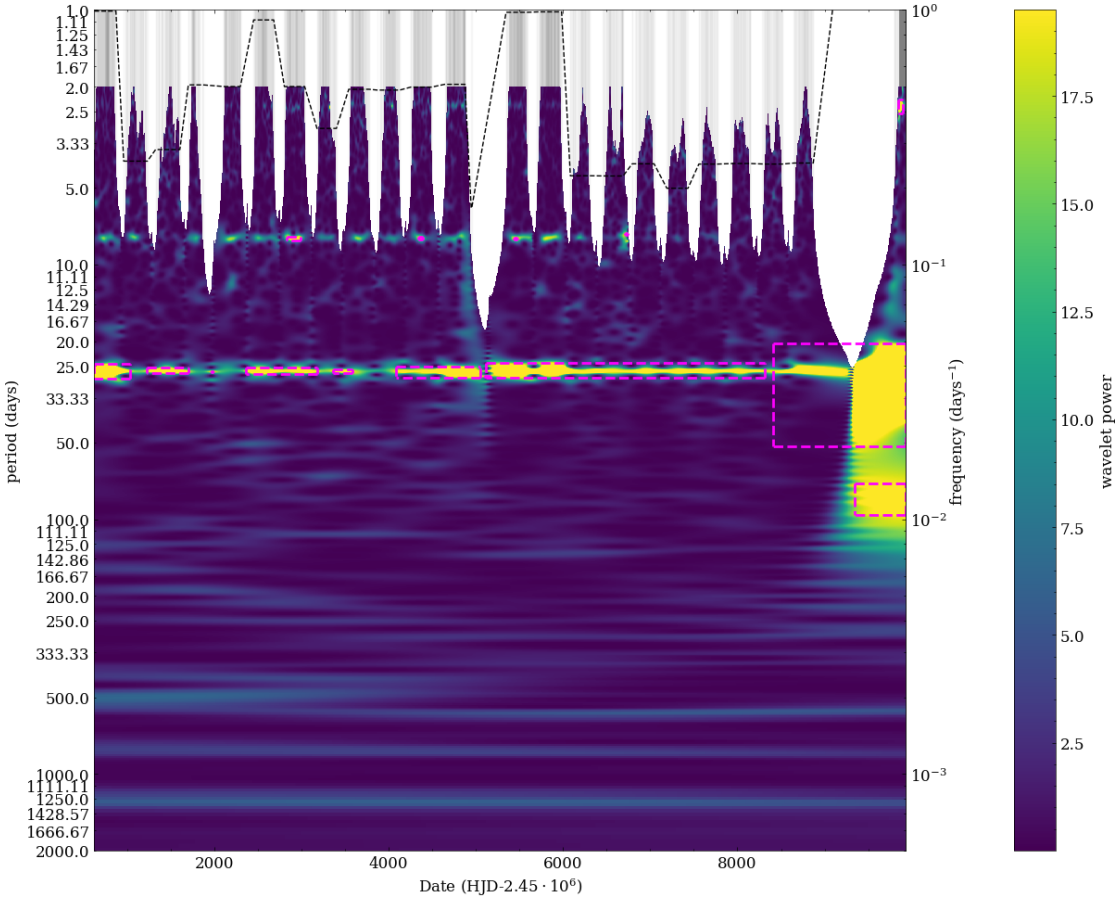


FIGURE A.30: WWZ transform of SXP1323 light curve.

References

J Aitchison, C Barceló-Vidal, J A Martín-Fernández, and V Pawlowsky-Glahn. Logratio Analysis and Compositional Distance. *Mathematical Geology*, 32(3):271–275, 2000. .

John Aitchison and Juan Jose Egozcue. Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology*, 37(7):829–850, 10 2005. ISSN 08828121. .

D. Altamirano, T. Belloni, M. Linares, M. van der Klis, R. Wijnands, P. A. Curran, M. Kalamkar, H. Stiele, S. Motta, T. Munoz-Darias, P. Casella, and H. Krimm. The faint "heartbeats" of IGR J17091-3624: an exceptional black-hole candidate. *The Astrophysical Journal Letters*, 742(2), 12 2011. . URL http://arxiv.org/abs/1112.2393http://dx.doi.org/10.1088/2041-8205/742/2/L17.

Mihael Ankerst, Markus M Breunig, Hans-peter Kriegel, and Jörg Sander. OPTICS: Ordering Points To Identify the Clustering Structure. *ACM SIGMOD Record*, 28(2):49–60, 1999.

D. J. Armstrong, J. Kirk, K. W. F. Lam, J. McCormac, H. P. Osborn, J. Spake, S. Walker, D. J. A. Brown, M. H. Kristiansen, D. Pollacco, R. West, and P. J. Wheatley. K2 Variable Catalogue II: Machine Learning Classification of Variable Stars and Eclipsing Binaries in K2 Fields 0-4. *Monthly Notices of the Royal Astronomical Society*, 456(2):2260–2272, 12 2015. . URL http://dx.doi.org/10.1093/mnras/stv2836.

Rob Ashmore, Radu Calinescu, and Colin Paterson. Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges. *ACM Computing Surveys*, 54(5), 6 2021. .

M Emre Aydin. eaydin/WWZ: First release, 3 2017. URL https://doi.org/10.5281/zenodo.375648.

T. Bagnoli and J. J.M. In't Zand. Discovery of GRS 1915+105 variability patterns in the Rapid Burster. *Monthly Notices of the Royal Astronomical Society: Letters*, 450(1):L52–L56, 4 2015. ISSN 17453933. .

Arash Bahramian and Nathalie Degenaar. Low-Mass X-ray Binaries. In *Handbook of X-ray and Gamma-ray Astrophysics*, pages 1–62. 6 2023. ISBN 978-981-16-4544-0. .

Luis A. Balona and Dogus Ozuyar. TESS Observations of Be Stars: General Characteristics and the Impulsive Magnetic Rotator Model. *The Astrophysical Journal*, 921(1):5, 11 2021. ISSN 0004-637X. .

R. V. Baluev. Assessing the statistical significance of periodogram peaks. *Monthly Notices of the Royal Astronomical Society*, 385(3):1279–1285, 2008. ISSN 13652966. .

Dalya Baron. Machine learning in astronomy: A practical overview. *arXiv preprint arXiv:1904.07248*, 2019.

Ardeshir Bayat. Science, medicine, and the future: Bioinformatics. *BMJ*, 324(7344):1018–1022, 2002. URL www.ebi.ac.uk.

I. Becker, K. Pichara, M. Catelan, P. Protopapas, C. Aguirre, and F. Nikzat. Scalable end-To-end recurrent neural network for variable star classification. *Monthly Notices of the Royal Astronomical Society*, 493(2):2981–2995, 4 2020. ISSN 13652966. .

Stefan Behnel, Robert Bradshaw, Craig Citro, Lisandro Dalcin, Dag Sverre Seljebotn, and Kurt Smith. Cython: The best of both worlds. *Computing in Science and Engineering*, 13(2): 31–39, 3 2011. ISSN 15219615. .

Eric C. Bellm. The Zwicky Transient Facility. In *The Third Hot-wiring the Transient Universe Workshop*, pages 27–33, 10 2014. URL http://arxiv.org/abs/1410.8185.

T Belloni, M Méndez, A R King, M Van Der Klis, and J Van Paradijs. AN UNSTABLE CENTRAL DISK IN THE SUPERLUMINAL BLACK HOLE X-RAY BINARY GRS 1915105. *The Astrophysical Journal Letters*, 479(2):145–148, 1997a. .

T Belloni, M Méndez, A R King, M Van Der Klis, and J Van Paradijs. A UNIFIED MODEL FOR THE SPECTRAL VARIABILITY IN GRS 1915+105. *The Astrophysical Journal Letters*, 488 (2):109–112, 1997b. .

T Belloni, M. Klein-Wolt, M. Mendez, M. van der Klis, and J. van Paradijs. A model-independent analysis of the variability of GRS 1915+105. *Astronomy and Astrophysics*, 355(1):271–290, 1 2000. URL http://arxiv.org/abs/astro-ph/0001103.

T. M. Belloni and D. Altamirano. High-frequency quasi-periodic oscillations from GRS 1915+105. *Monthly Notices of the Royal Astronomical Society*, 432(1):10–18, 2013. ISSN 00358711. .

T M Belloni, S E Motta, and T Muñoz-Darias. Black hole transients. *Bull. Astr. Soc. India*, 39 (3):409–428, 2011.

Tomaso Belloni. GRS 1915+105 as a black hole accretion disk laboratory. In *The Neutron Star-Black Hole Connection*, volume 567, pages 295–300. Kluwer Academic Publishers, 2001. ISBN 978-1-4020-0204-5.

Seif-Eddine Benkabou, Khalid Benabdeslem, and Bruno Canitia. Unsupervised outlier detection for time series by entropy and dynamic time warping. *Knowledge and Information Systems*, 54(2):463–486, 2 2018. ISSN 0219-1377. . URL http://link.springer.com/10.1007/s10115-017-1067-8.

Arnold O. Benz and Manuel Güdel. Physical processes in magnetically driven flares on the sun, stars, young stellar objects. *Annual Review of Astronomy and Astrophysics*, 48:241–287, 9 2010. .

A. J. Bird, M. J. Coe, V. A. Mcbride, and A. Udalski. On the periodicities present in the optical light curves of Small Magellanic Cloud Be/X-ray binaries. *Monthly Notices of the Royal Astronomical Society*, 423(4):3663–3678, 2012. ISSN 00358711. .

Leo Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

David N. Burrows, J. E. Hill, J. A. Nousek, J. A. Kennea, A. Wells, J. P. Osborne, A. F. Abbey, A. Beardmore, K. Mukerjee, A. D.T. Short, G. Chincarini, S. Campana, O. Citterio, A. Moretti, C. Pagani, G. Tagliaferri, P. Giommi, M. Capalbi, F. Tamburelli, L. Angelini, G. Cusumano, H. W. Bräuninger, W. Burkert, and G. D. Hartner. The Swift X-ray telescope. *Space Science Reviews*, 120:165–195, 12 2005. ISSN 00386308. .

Richard H Byrdt, Peihuang Lut, Jorge Nocedalt, and Ciyou Zhu. A LIMITED MEMORY ALGORITHM FOR BOUND CONSTRAINED OPTIMIZATION. *SIAM J. ScI. COMPUT*, 16(5):1190–1208, 1995. URL https://epubs.siam.org/terms-privacy.

F Capitanio, A Bazzano, P Ubertini, A A Zdziarski, A J Bird, G De Cesare, A J Dean, J B Stephen, and A Tarana. Spectral States of the X-Ray Binary IGR J17091-3624 Observed by INTEGRAL and RXTE. *The Astrophysical Journal*, 643(1):376–380, 2006.

S. Carpano, F. Haberl, and R. Sturm. Discovery of a 26.2 day period in the long-term X-ray light curve of SXP 1323: A very short orbital period for a long spin period pulsar. *Astronomy and Astrophysics*, 602, 6 2017. ISSN 14320746. .

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017. ISSN 15487660. .

A. Castro, D. Altamirano, R. Michel, P. Gandhi, J. V.Hernández Santisteban, J. Echevarría, C. Tejada, C. Knigge, G. Sierra, E. Colorado, J. Hernández-Landa, D. Whiter, M. Middleton, B. García, G. Guisa, and N. Castro Segura. Opticam: A triple-camera optical system designed to explore the fastest timescales in astronomy. *Revista Mexicana de Astronomia y Astrofisica*, 55(2):363–376, 2019. ISSN 01851101. .

Alberto J Castro-Tirado, Soren Brandt, Niels Lund, Igor Lapshov, Rashid A. Sunyaev, Aleksei A. Shlyapnikov, Sergei Guziy, and Elena P. Pavlenko. Discovery and Observations by WATCH of the X-Ray Transient GRS 1915+105. *Astrophysical Journal Supplement*, 92:469–472, 1994. .

M. Catelan and H. A. Smith. *Pulsating Stars*. John Wiley & Sons, 2015. ISBN 9783527655182. .

Tom Charnock and Adam Moss. Deep Recurrent Neural Networks for Supernovae Classification. *The Astrophysical Journal*, 837(2):L28, 3 2017. ISSN 1538-4357. .

Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. cuDNN: Efficient Primitives for Deep Learning, 10 2014. URL http://arxiv.org/abs/1410.0759.

François Chollet. Keras, 2015. URL https://keras.io.

W I Clarkson, P A Charles, M J Coe, S Laycock, M D Tout, and C A Wilson. Long-term properties of accretion discs in X-ray binaries-I. The variable third period in SMC X-1. *Mon. Not. R. Astron. Soc*, 339:447–454, 2003. URL https://academic.oup.com/mnras/article/339/2/447/1005756.

M. J. Coe and J. Kirk. Catalogue of Be/X-ray binary systems in the Small Magellanic Cloud: X-ray, optical and IR properties. *Monthly Notices of the Royal Astronomical Society*, 452(1): 969–977, 4 2015. ISSN 13652966. .

M J Coe and J A Orosz. OGLE observations of four X-ray binary pulsars in the Small Magellanic Cloud. *Mon. Not. R. Astron. Soc*, 311:169–175, 2000. URL https://academic.oup.com/mnras/article/311/1/169/990379.

M. J. Coe, W. R.T. Edge, J. L. Galache, and V. A. McBride. Optical properties of Small Magellanic Cloud X-ray binaries. *Monthly Notices of the Royal Astronomical Society*, 356(2): 502–514, 1 2005. ISSN 00358711. .

M. J. Coe, M. Schurch, R. H.D. Corbet, J. Galache, V. A. McBride, L. J. Townsend, and A. Udalski. An optical and X-ray study of the counterpart to the Small Magellanic Cloud X-ray binary pulsar system SXP327. *Monthly Notices of the Royal Astronomical Society*, 387 (2):724–728, 2008. ISSN 13652966. .

R. H. D. Corbet. The three types of high-mass X-ray pulsator. *Monthly Notices of the Royal Astronomical Society*, 220(1):1047–1056, 1986.

J. M.C. Court, D. Altamirano, M. Pereyra, C. M. Boon, K. Yamaoka, T. Belloni, R. Wijnands, and M. Pahari. An atlas of exotic variability in IGR J17091-3624: A comparison with GRS 1915+105. *Monthly Notices of the Royal Astronomical Society*, 468(4):4748–4771, 7 2017. ISSN 13652966. .

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

N. Degenaar, M. T. Wolff, P. S. Ray, K. S. Wood, J. Homan, W. H.G. Lewin, P. G. Jonker, E. M. Cackett, J. M. Miller, E. F. Brown, and R. Wijnands. Further X-ray observations of EXO 0748-676 in quiescence: Evidence for a cooling neutron star crust. *Monthly Notices of the Royal Astronomical Society*, 412(3):1409–1418, 2011. ISSN 13652966. .

A P Dempster, N M Laird, and D B Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

V. S. Dhillon, T. R. Marsh, M. J. Stevenson, D. C. Atkinson, P. Kerry, P. T. Peacocke, A. J. A. Vick, S. M. Beard, D. J. Ives, D. W. Lunney, S. A. McLay, C. J. Tierney, J. Kelly, S. P. Littlefair, R. Nicholson, R. Pashley, E. T. Harlaftis, and K. O'Brien. ULTRACAM: an ultra-fast, triple-beam CCD camera for high-speed astrophysics. *Monthly Notices of the Royal Astronomical Society*, 378(3):825–840, 4 2007. . URL http://arxiv.org/abs/0704.2557http://dx.doi.org/10.1111/j.1365-2966.2007.11881.x.

V. S. Dhillon, T. R. Marsh, N. Bezawada, M. Black, S. Dixon, T. Gamble, D. Henry, P. Kerry, S. P. Littlefair, D. W. Lunney, T. Morris, J. Osborn, and R. W. Wilson. HiPERCAM: A high-speed, quintuple-beam CCD camera for the study of rapid variability in the Universe. In *Proceedings of the SPIE*, page 99080Y, 6 2016. . URL http://arxiv.org/abs/1606.09214http://dx.doi.org/10.1117/12.2229055.

Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

Chris Done, Grzegorz Wardziński, and Marek Gierliński. GRS 1915+105: The brightest Galactic black hole. *Monthly Notices of the Royal Astronomical Society*, 349(2):393–403, 4 2004. ISSN 00358711. .

Wei Dong, Charikar Moses, and Kai Li. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World wide web*, pages 577–586, 2011.

W. R.T. Edge, M. J. Coe, J. L. Galache, V. A. McBride, R. H.D. Corbet, C. B. Markwardt, and S. Laycock. Three new X-ray pulsars detected in the Small Magellanic Cloud and the positions of two other known pulsars determined. *Monthly Notices of the Royal Astronomical Society*, 353(4):1286–1292, 10 2004. ISSN 00358711. .

W. R.T. Edge, M. J. Coe, J. L. Galache, V. A. McBride, R. H.D. Corbet, A. T. Okazaki, S. Laycock, C. B. Markwardt, F. E. Marshall, and A. Udalski. The binary period and outburst behaviour of the Small Magellanic Cloud X-ray binary pulsar system SXP504. *Monthly Notices of the Royal Astronomical Society*, 361(2):743–747, 8 2005. ISSN 00358711. .

William Raymond Tarratt Edge. *Studies of Be X-ray binaries in the Magellanic Clouds*. PhD thesis, University of Southampton, 2005. URL https://eprints.soton.ac.uk/465791/.

P. Eger and F. Haberl. XMM-Newton observations of the small magellanic cloud: RXJ0105.9-7203, a 726 s Be/X-ray binary pulsar. *Astronomy and Astrophysics*, 485(3):807–811, 2008a. ISSN 14320746. .

P. Eger and F. Haberl. XMM-Newton observations of the Small Magellanic Cloud: Long term evolution of frequently observed Be/X-ray binaries. *Astronomy and Astrophysics*, 491(3): 841–849, 12 2008b. ISSN 00046361. .

Ronen Eldan and Ohad Shamir. The power of depth for feedforward neural networks. In *Conference on learning theory*, pages 907–940, 2016.

Martin Ester, Hans-Peter Kriegel, Jiirg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, 1996. URL www.aaai.org.

L. Eyer, L. Rimoldini, M. Audard, R. I. Anderson, K. Nienartowicz, F. Glass, O. Marchal, M. Grenon, N. Mowlavi, B. Holl, G. Clementini, C. Aerts, T. Mazeh, D. W. Evans, L. Szabados, A. G.A. Brown, A. Vallenari, T. Prusti, J. H.J. De Bruijne, C. Babusiaux, C. A.L. Bailer-Jones, M. Biermann, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, H. I. Siddiqui, C. Soubiran, F. Van Leeuwen, N. A. Walton, F. Arenou, U. Bastian, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, J. Bakker, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, E. Masana, R. Messineo, P. Panuzzo, J. Portell, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, D. Teyssier, M. Altmann, R. Andrae, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. Burgess, G. Busso, B. Carry, A. Cellino, M. Clotet, O. Creevey, M. Davidson, J. De Ridder, L. Delchambre, A. Dell'Oro, C. Ducourant, J. Fernández-Hernández, M. Fouesneau, Y. Frémat, L. Galluccio, M. García-Torres, J. González-Núñez, J. J. González-Vidal, E. Gosset, L. P. Guy, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, A. Hutton, G. Jasniewicz, A. Jean-Antoine-Piccolo, S. Jordan, A. J. Korn, A. Krone-Martins, A. C. Lanzafame, T. Lebzelter, W. Löffler, M. Manteiga, P. M. Marrese, J. M. Martín-Fleitas, A. Moitinho, A. Mora, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J. M. Petit, A. Recio-Blanco, P. J. Richards, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, M. Süveges, J. Torra, W. Van Reeven, U. Abbas, A. Abreu Aramburu, S. Accart, G. Altavilla, M. A. Álvarez, R. Alvarez, J. Alves, A. H. Andrei, E. Anglada Varela, E. Antiche, T. Antoja, B. Arcay, T. L. Astraatmadja, N. Bach, S. G. Baker, L. Balaguer-Núñez, P. Balm, C. Barache, C. Barata, D. Barbato, F. Barblan, P. S. Barklem, D. Barrado, M. Barros, M. A. Barstow, S. Bartholomé Muñoz, J. L. Bassilana, U. Becciani, M. Bellazzini, A. Berihuete, S. Bertone, L. Bianchi, O. Bienaymé, S. Blanco-Cuaresma, T. Boch, C. Boeche, A. Bombrun, R. Borrachero, D. Bossini, S. Bouquillon, G. Bourda, A. Bragaglia, L. Bramante, M. A. Breddels, A. Bressan, N. Brouillet, T. Brüsemeister, E. Brugaletta, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, G. Cannizzaro, T. Cantat-Gaudin, R. Carballo, T. Carlucci, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, P. Charlot, L. Chemin, A. Chiavassa, G. Cocozza, G. Costigan, S. Cowell, F. Crifo, M. Crosta, C. Crowley, J. Cuypers, C. Dafonte, Y. Damerdji, A. Dapergolas, P. David, M. David, P. De Laverny, F. De Luise, R. De March, D. De Martino, R. De Souza, A. De Torres, J. Debosscher, E. Del Pozo, M. Delbo, A. Delgado, H. E. Delgado, S. Diakite, C. Diener, E. Distefano, C. Dolding, P. Drazinos, J. Durán, B. Edvardsson, H. Enke, K. Eriksson, P. Esquej, G. Eynard Bontemps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão,

M. Farràs Casas, L. Federici, G. Fedorets, P. Fernique, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, E. Fraile, M. Fraser, B. Frézouls, M. Gai, S. Galleti, D. Garabato, F. García-Sedano, A. Garofalo, N. Garralda, A. Gavel, P. Gavras, J. Gerssen, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, M. Gomes, M. Granvik, A. Gueguen, A. Guerrier, J. Guiraud, R. Gutiérrez-Sánchez, R. Haigron, D. Hatzidimitriou, M. Hauser, M. Haywood, U. Heiter, A. Helmi, J. Heu, T. Hilger, D. Hobbs, W. Hofmann, G. Holland, H. E. Huckle, A. Hypki, V. Icardi, K. Janßen, G. Jevardat De Fombelle, P. G. Jonker, A. L. Juhász, F. Julbe, A. Karampelas, A. Kewley, J. Klar, A. Kochoska, R. Kohley, K. Kolenberg, M. Kontizas, E. Kontizas, S. E. Koposov, G. Kordopatis, Z. Kostrzewa-Rutkowska, P. Koubsky, S. Lambert, A. F. Lanza, Y. Lasne, J. B. Lavigne, Y. Le Fustec, C. Le Poncin-Lafitte, Y. Lebreton, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, H. Lenhardt, F. Leroux, S. Liao, E. Licata, H. E.P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, M. López, D. Lorenz, S. Managau, R. G. Mann, G. Mantelet, J. M. Marchant, M. Marconi, S. Marinoni, G. Marschalkó, D. J. Marshall, M. Martino, G. Marton, N. Mary, D. Massari, G. Matijevič, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, D. Molina, R. Molinaro, L. Molnár, P. Montegriffo, R. Mor, R. Morbidelli, T. Morel, S. Morgenthaler, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, G. Nelemans, L. Nicastro, L. Noval, W. O'Mullane, C. Ordénovic, D. Ordóñez-Blanco, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, A. Panahi, M. Pawlak, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, E. Poujoulet, A. Prša, L. Pulone, E. Racero, S. Ragaini, N. Rambaux, M. Ramos-Lerate, S. Regibo, C. Reylé, F. Riclet, V. Ripepi, A. Riva, A. Rivard, G. Rixon, T. Roegiers, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, N. Sanna, T. Santana-Ros, M. Sarasso, H. Savietto, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, I. C. Shih, L. Siltala, A. F. Silva, R. L. Smart, K. W. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay, A. Spagna, F. Spoto, U. Stampa, I. A. Steele, H. Steidelmüller, C. A. Stephenson, H. Stoev, F. F. Suess, J. Surdej, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrett, P. Teyssandier, W. Thuillot, A. Titarenko, F. Torra Clotet, C. Turon, A. Ulla, E. Utrilla, S. Uzzi, M. Vaillant, G. Valentini, V. Valette, A. Van Elteren, E. Van Hemelryck, M. Van Leeuwen, M. Vaschetto, A. Vecchiato, J. Veljanoski, Y. Viala, D. Vicente, S. Vogt, C. Von Essen, H. Voss, V. Votruba, S. Voutsinas, G. Walmsley, M. Weiler, O. Wertz, T. Wevers, A. Wyrzykowski, A. Yoldas, M. Žerjal, H. Ziaeepour, J. Zorec, S. Zschocke, S. Zucker, C. Zurbach, and T. Zwitter. Gaia Data Release 2: Variable stars in the colour-absolute magnitude diagram. *Astronomy and Astrophysics*, 623: 1–20, 3 2019. .

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

Rob Fender and Tomaso Belloni. GRS 1915+105 and the Disc-Jet Coupling in Accreting Black Hole Systems. *Annual Review of Astronomy and Astrophysics*, 42(1):317–364, 9 2004. ISSN 0066-4146. . URL http://www.annualreviews.org/doi/10.1146/annurev.astro.42.053102.134031.

Grant Foster. Wavelets for period analysis of unevenly sampled time series. *Astronomical Journal*, 112(4):1709–1729, 1996.

Juhan. Frank, Andrew. King, and Derek. Raine. Accretion Power in Astrophysics: Third Edition, 2002. ISSN 00319228. URL http://adsabs.harvard.edu/abs/2002apa..book.....F.

Dennis Gabor. Theory of Communication. *Journal of Institution of Electrical Engineers*, 93(3): 429–457, 1946.

J L Galache, R H D Corbet, M J Coe, S Laycock, M P E Schurch, C Markwardt, F E Marshall, and J Lochner. A LONG LOOK AT THE Be/X-RAY BINARIES OF THE SMALL MAGELLANIC CLOUD. *The Astrophysical Journal Supplement Series*, 177(1):189–215, 2008.

Milton A. Garcés, Daniel Bowman, Cleat Zeiler, Anthony Christe, Tyler Yoshiyama, Brian Williams, Meritxell Colet, Samuel Takazawa, and Sarah Popenhagen. Skyfall: Signal Fusion from a Smartphone Falling from the Stratosphere. *Signals*, 3(2):209–234, 4 2022. .

N Gehrels, G Chincarini, P Giommi, K O Mason, J A Nousek, A A Wells, N E White, S D Barthelmy, D N Burrows, L R Cominsky, K C Hurley, F E Marshall, P Mészáros, P W A Roming, L Angelini, L M Barbier, T Belloni, S Campana, P A Caraveo, M M Chester, O Citterio, T L Cline, M S Cropper, J R Cummings, A J Dean, E D Feigelson, E E Fenimore, D A Frail, A S Fruchter, G P Garmire, K Gendreau, G Ghisellini, J Greiner, J E Hill, S D Hunsberger, H A Krimm, S R Kulkarni, P Kumar, F Lebrun, N M Lloyd-Ronning, C B Markwardt, B J Mattson, R F Mushotzky, J P Norris, J Osborne, B Paczynski, D M Palmer, H.-S Park, A M Parsons, J Paul, M J Rees, C S Reynolds, J E Rhoads, T P Sasseen, B E Schaefer, A T Short, A P Smale, I A Smith, L Stella, G Tagliaferri, T Takahashi, M Tashiro, L K Townsley, J Tueller, M J L Turner, M Vietri, W Voges, M J Ward, R Willingale, F M Zerbi, and W W Zhang. THE SWIFT GAMMA-RAY BURST MISSION. *The Astrophysical Journal*, 611(2):1005–1020, 2004.

Keith C. Gendreau, Zaven Arzoumanian, Phillip W. Adkins, Cheryl L. Albert, John F. Anders, Andrew T. Aylward, Charles L. Baker, Erin R. Balsamo, William A. Bamford, Suyog S. Benegalrao, Daniel L. Berry, Shiraz Bhalwani, J. Kevin Black, Carl Blaurock, Ginger M. Bronke, Gary L. Brown, Jason G. Budinoff, Jeffrey D. Cantwell, Thoniel Cazeau, Philip T. Chen, Thomas G. Clement, Andrew T. Colangelo, Jerry S. Coleman, Jonathan D. Coopersmith, William E. Dehaven, John P. Doty, Mark D. Egan, Teruaki Enoto, Terry W. Fan, Deneen M. Ferro, Richard Foster, Nicholas M. Galassi, Luis D. Gallo, Chris M. Green, Dave Grosh, Kong Q. Ha, Monther A. Hasouneh, Kristofer B. Heefner, Phyllis Hestnes, Lisa J. Hoge, Tawanda M. Jacobs, John L. Jørgensen, Michael A. Kaiser, James W. Kellogg, Steven J. Kenyon, Richard G. Koenecke, Robert P. Kozon, Beverly LaMarr, Mike D. Lambertson, Anne M. Larson, Steven Lentine, Jesse H. Lewis, Michael G. Lilly, Kuochia Alice Liu, Andrew Malonis, Sridhar S. Manthripragada, Craig B. Markwardt, Bryan D. Matonak, Isaac E. Mcginnis, Roger L. Miller, Alissa L. Mitchell, Jason W. Mitchell, Jelila S. Mohammed,

Charles A. Monroe, Kristina M. Montt de Garcia, Peter D. Mulé, Louis T. Nagao, Son N. Ngo, Eric D. Norris, Dwight A. Norwood, Joseph Novotka, Takashi Okajima, Lawrence G. Olsen, Chimaobi O. Onyeachu, Henry Y. Orosco, Jacqualine R. Peterson, Kristina N. Pevear, Karen K. Pham, Sue E. Pollard, John S. Pope, Daniel F. Powers, Charles E. Powers, Samuel R. Price, Gregory Y. Prigozhin, Julian B. Ramirez, Winston J. Reid, Ronald A. Remillard, Eric M. Rogstad, Glenn P. Rosecrans, John N. Rowe, Jennifer A. Sager, Claude A. Sanders, Bruce Savadkin, Maxine R. Saylor, Alexander F. Schaeffer, Nancy S. Schweiss, Sean R. Semper, Peter J. Serlemitsos, Larry V. Shackelford, Yang Soong, Jonathan Struebel, Michael L. Vezie, Joel S. Villasenor, Luke B. Winternitz, George I. Wofford, Michael R. Wright, Mike Y. Yang, and Wayne H. Yu. The Neutron star Interior Composition Explorer (NICER): design and development. In *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*, volume 9905, page 99051H. SPIE, 7 2016. ISBN 9781510601895. .

M. Gilfanov. X-Ray Emission from Black-Hole Binaries. In *The Jet Paradigm: From Microquasars to Quasars*, pages 17–51. 2010. .

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL http://www.deeplearningbook.org.

D. E. Gruber, P. R. Blanco, W. A. Heindl, M. R. Pelling, R. E. Rothschild, and P. L. Hink. The high energy X-ray timing experiment on XTE. *Astronomy and Astrophysics Supplement Series*, 120:641–644, 1996.

F. Haberl, C. Maitra, S. Carpano, X. Dai, V. Doroshenko, K. Dennerl, M. J. Freyberg, M. Sasaki, A. Udalski, K. A. Postnov, and N. I. Shakura. eROSITA calibration and performance verification phase: High-mass X-ray binaries in the Magellanic Clouds. *Astronomy and Astrophysics*, 661, 5 2022. ISSN 14320746. .

D. C. Hannikainen, O. Vilhu, J. Rodriguez, S. Brandt, N. J. Westergaard, N. Lund, I. Mocœur, Ph. Durouchoux, T. Belloni, A. Castro-Tirado, P. A. Charles, A. J. Dean, R. P. Fender, M. Feroci, P. Hakala, R. W. Hunstead, C. R. Kaiser, A. King, I. F. Mirabel, G. G. Pooley, J. Poutanen, K. Wu, and A. A. Zdziarski. First INTEGRAL observations of GRS 1915+105. *Astronomy & Astrophysics*, 411(1):415–419, 11 2003. ISSN 0004-6361. . URL http://www.aanda.org/10.1051/0004-6361:20031444.

D. C. Hannikainen, J. Rodriguez, O. Vilhu, L. Hjalmarsdotter, A. A. Zdziarski, T. Belloni, J. Poutanen, K. Wu, S. E. Shaw, V. Beckmann, R. W. Hunstead, G. G. Pooley, N. J. Westergaard, I. F. Mirabel, P. Hakala, A. Castro-Tirado, and Ph. Durouchoux. Characterizing a new class of variability in GRS 1915+105 with simultaneous INTEGRAL/RXTE observations. *Astronomy & Astrophysics*, 435(3):995–1004, 6 2005. ISSN 0004-6361. . URL http://www.aanda.org/10.1051/0004-6361:20042250.

K. P. Harikrishnan, Ranjeev Misra, and G. Ambika. Nonlinear time series analysis of the light curves from the black hole system GRS1915+105. *Research in Astronomy and Astrophysics*, 11(1):71–90, 2011. ISSN 16744527. .

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, 9 2020. ISSN 14764687. .

J. D. Hartman and G. Bakos. Vartools: A program for analyzing astronomical time-series data. *Astronomy and Computing*, 17:1–72, 10 2016. ISSN 22131337. .

N. Heckert and J. Filliben. NIST/SEMATECH e-Handbook of Statistical Methods. In *NIST/SEMATECH e-Handbook of Statistical Methods*, chapter 1. 2003. URL http://www.itl.nist.gov/div898/handbook/.

L. M. Heil, P. Uttley, and M. Klein-Wolt. Inclination-dependent spectral and timing properties in transient black hole X-ray binaries. *Monthly Notices of the Royal Astronomical Society*, 448 (4):3348–3353, 4 2015. ISSN 13652966. .

A. N. Heinze, J. L. Tonry, L. Denneau, H. Flewelling, B. Stalder, A. Rest, K. W. Smith, S. J. Smartt, and H. Weiland. A first catalog of variable stars measured by the asteroid terrestrial-impact last alert system (ATLAS). *The Astronomical Journal*, 156(5), 4 2018. ISSN 23318422. .

J. V. Hernández Santisteban, V. Cúneo, N. Degenaar, J. van den Eijnden, D. Altamirano, M. N. Gómez, D. M. Russell, R. Wijnands, R. Golovakova, M. T. Reynolds, and J. M. Miller. Multiwavelength characterization of the accreting millisecond X-ray pulsar and ultracompact binary IGR J17062-6143. *Monthly Notices of the Royal Astronomical Society*, 488(4):4596–4606, 10 2019. ISSN 13652966. .

Geoffrey Hinton. Non-linear dimensionality reduction, 2013. URL https://www.cs.toronto.edu/~hinton/csc2535/notes/lec11new.pdf.

Sepp Hochreiter and Jurgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9 (8):1735–1780, 1997. .

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991. ISSN 0893-6080. . URL https://www.sciencedirect.com/science/article/pii/089360809190009T.

Daniela Huppenkothen, Lucy M. Heil, David W. Hogg, and Andreas Mueller. Using machine learning to explore the long-term evolution of GRS 1915+105. *Monthly Notices of the Royal Astronomical Society*, 466(2):2364–2377, 4 2017. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw3190.

Vojtech Huser and James J. Cimino. Impending Challenges for the Use of Big Data. *International Journal of Radiation Oncology Biology Physics*, 95(3):890–894, 7 2016. .

Rob J. Hyndman, Earo Wang, and Nikolay Laptev. Large-Scale Unusual Time Series Detection. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1616–1619, 2015. ISBN 9781467384926. .

Robert I. Hynes. Multiwavelength observations of accretion in low-mass X-ray binary systems. In *Accretion Processes In Astrophysics: XXI Canary Islands Winter School Of Astrophysics*, volume 9781107030190, pages 117–150. Cambridge University Press, 1 2012. ISBN 9781139343268. .

Andrei P. Igoshev, Sergei B. Popov, and Rainer Hollerbach. Evolution of neutron star magnetic fields. *Universe*, 7(9), 9 2021. ISSN 22181997. .

Adam R. Ingram and Sara E. Motta. A review of quasi-periodic oscillations from black hole X-ray binaries: Observation and theory. *New Astronomy Reviews*, 85:101524, 9 2019. ISSN 13876473. .

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019. .

Zeljko Ivezic, Steven M. Kahn, J. Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F. Anderson, John Andrew, James Roger P. Angel, George Z. Angeli, Reza Ansari, Pierre Antilogus, Constanza Araujo, Robert Armstrong, Kirk T. Arndt, Pierre Astier, Éric Aubourg, Nicole Auza, Tim S. Axelrod, Deborah J. Bard, Jeff D. Barr, Aurelian Barrau, James G. Bartlett, Amanda E. Bauer, Brian J. Bauman, Sylvain Baumont, Ellen Bechtol, Keith Bechtol, Andrew C. Becker, Jacek Becla, Cristina Beldica, Steve Bellavia, Federica B. Bianco, Rahul Biswas, Guillaume Blanc, Jonathan Blazek, Roger D. Blandford, Josh S. Bloom, Joanne Bogart, Tim W. Bond, Michael T. Booth, Anders W. Borgland, Kirk Borne, James F. Bosch, Dominique Boutigny, Craig A. Brackett, Andrew Bradshaw, William Nielsen Brandt, Michael E. Brown, James S. Bullock, Patricia Burchat, David L. Burke, Gianpietro Cagnoli, Daniel Calabrese, Shawn Callahan, Alice L. Callen, Jeffrey L. Carlin, Erin L. Carlson, Srinivasan Chandrasekharan, Glenaver Charles-Emerson, Steve Chesley, Elliott C. Cheu, Hsin-Fang Chiang, James Chiang, Carol Chirino, Derek Chow, David R. Ciardi, Charles F. Claver, Johann Cohen-Tanugi, Joseph J. Cockrum, Rebecca Coles, Andrew J. Connolly, Kem H. Cook, Asantha Cooray, Kevin R. Covey, Chris Cribbs, Wei Cui, Roc Cutri, Philip N. Daly, Scott F. Daniel, Felipe Daruich, Guillaume Daubard, Greg Daues, William Dawson, Francisco Delgado, Alfred Dellapenna, Robert de Peyster, Miguel de Val-Borro, Seth W. Digel, Peter Doherty, Richard Dubois, Gregory P. Dubois-Felsmann, Josef Durech, Frossie Economou, Tim Eifler, Michael Eracleous, Benjamin L. Emmons, Angelo Fausti Neto, Henry Ferguson, Enrique Figueroa, Merlin Fisher-Levine, Warren Focke, Michael D. Foss, James Frank, Michael D. Freemon, Emmanuel Gangler, Eric Gawiser, John C. Geary, Perry Gee, Marla Geha, Charles J. B. Gessner, Robert R. Gibson, D. Kirk Gilmore, Thomas Glanzman, William Glick, Tatiana Goldina, Daniel A. Goldstein, Iain Goodenow, Melissa L. Graham, William J. Gressler, Philippe Gris, Leanne P.

Guy, Augustin Guyonnet, Gunther Haller, Ron Harris, Patrick A. Hascall, Justine Haupt, Fabio Hernandez, Sven Herrmann, Edward Hileman, Joshua Hoblitt, John A. Hodgson, Craig Hogan, James D. Howard, Dajun Huang, Michael E. Huffer, Patrick Ingraham, Walter R. Innes, Suzanne H. Jacoby, Bhuvnesh Jain, Fabrice Jammes, James Jee, Tim Jenness, Garrett Jernigan, Darko Jevremović, Kenneth Johns, Anthony S. Johnson, Margaret W. G. Johnson, R. Lynne Jones, Claire Juramy-Gilles, Mario Jurić, Jason S. Kalirai, Nitya J. Kallivayalil, Bryce Kalmbach, Jeffrey P. Kantor, Pierre Karst, Mansi M. Kasliwal, Heather Kelly, Richard Kessler, Veronica Kinnison, David Kirkby, Lloyd Knox, Ivan V. Kotov, Victor L. Krabbendam, K. Simon Krughoff, Petr Kubánek, John Kuczewski, Shri Kulkarni, John Ku, Nadine R. Kurita, Craig S. Lage, Ron Lambert, Travis Lange, J. Brian Langton, Laurent Le Guillou, Deborah Levine, Ming Liang, Kian-Tat Lim, Chris J. Lintott, Kevin E. Long, Margaux Lopez, Paul J. Lotz, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ashish Mahabal, Rachel Mandelbaum, Thomas W. Markiewicz, Darren S. Marsh, Philip J. Marshall, Stuart Marshall, Morgan May, Robert McKercher, Michelle McQueen, Joshua Meyers, Myriam Migliore, Michelle Miller, David J. Mills, Connor Miraval, Joachim Moeyens, Fred E. Moolekamp, David G. Monet, Marc Moniez, Serge Monkewitz, Christopher Montgomery, Christopher B. Morrison, Fritz Mueller, Gary P. Muller, Freddy Muñoz Arancibia, Douglas R. Neill, Scott P. Newbry, Jean-Yves Nief, Andrei Nomerotski, Martin Nordby, Paul O'Connor, John Oliver, Scot S. Olivier, Knut Olsen, William O'Mullane, Sandra Ortiz, Shawn Osier, Russell E. Owen, Reynald Pain, Paul E. Palecek, John K. Parejko, James B. Parsons, Nathan M. Pease, J. Matt Peterson, John R. Peterson, Donald L. Petravick, M. E. Libby Petrick, Cathy E. Petry, Francesco Pierfederici, Stephen Pietrowicz, Rob Pike, Philip A. Pinto, Raymond Plante, Stephen Plate, Joel P. Plutchak, Paul A. Price, Michael Prouza, Veljko Radeka, Jayadev Rajagopal, Andrew P. Rasmussen, Nicolas Regnault, Kevin A. Reil, David J. Reiss, Michael A. Reuter, Stephen T. Ridgway, Vincent J. Riot, Steve Ritz, Sean Robinson, William Roby, Aaron Roodman, Wayne Rosing, Cecille Roucelle, Matthew R. Rumore, Stefano Russo, Abhijit Saha, Benoit Sassolas, Terry L. Schalk, Pim Schellart, Rafe H. Schindler, Samuel Schmidt, Donald P. Schneider, Michael D. Schneider, William Schoening, German Schumacher, Megan E. Schwamb, Jacques Sebag, Brian Selvy, Glenn H. Sembroski, Lynn G. Seppala, Andrew Serio, Eduardo Serrano, Richard A. Shaw, Ian Shipsey, Jonathan Sick, Nicole Silvestri, Colin T. Slater, J. Allyn Smith, R. Chris Smith, Shahram Sobhani, Christine Soldahl, Lisa Storrie-Lombardi, Edward Stover, Michael A. Strauss, Rachel A. Street, Christopher W. Stubbs, Ian S. Sullivan, Donald Sweeney, John D. Swinbank, Alexander Szalay, Peter Takacs, Stephen A. Tether, Jon J. Thaler, John Gregg Thayer, Sandrine Thomas, Adam J. Thornton, Vaikunth Thukral, Jeffrey Tice, David E. Trilling, Max Turri, Richard Van Berg, Daniel Vanden Berk, Kurt Vetter, Francoise Virieux, Tomislav Vucina, William Wahl, Lucianne Walkowicz, Brian Walsh, Christopher W. Walter, Daniel L. Wang, Shin-Yawn Wang, Michael Warner, Oliver Wiecha, Beth Willman, Scott E. Winters, David Wittman, Sidney C. Wolff, W. Michael Wood-Vasey, Xiuqin Wu, Bo Xin, Peter Yoachim, and Hu Zhan. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873(2):111, 3 2019. ISSN 1538-4357.

.

Keith Jahoda, Jean H. Swank, Alan B. Giles, Michael J. Stark, Tod Strohmayer, William W. Zhang, and Edward H. Morgan. In-orbit performance and calibration of the Rossi X-ray Timing Explorer (RXTE) Proportional Counter Array (PCA). In *EUV, X-Ray, and Gamma-Ray Instrumentation for Astronomy VII*, volume 2808, pages 59–70. SPIE, 10 1996. ISBN 9780819421968. .

Shruti Kaushik, Abhinav Choudhury, Pankaj Kumar Sheron, Nataraj Dasgupta, Sayee Natarajan, Larry A. Pickett, and Varun Dutt. AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Frontiers in Big Data*, 3, 3 2020. ISSN 2624909X. .

J. A. Kennea, M. J. Coe, and P. A. Evans. SXP 304: S-CUBED detection of an outburst, possible orbital period. *The Astronomer's Telegram*, 12236:1, 11 2018a.

J. A. Kennea, M. J. Coe, P. A. Evans, J. Waters, and R. E. Jasko. The First Year of S-CUBED: The Swift Small Magellanic Cloud Survey . *The Astrophysical Journal*, 868(1):47, 11 2018b. ISSN 15384357. .

Curtis E. Kennedy and James P. Turley. Time series analysis as input for clinical predictive modeling: Modeling cardiac arrest in a pediatric ICU. *Theoretical Biology and Medical Modelling*, 8(1), 2011. ISSN 17424682. .

Jeffrey O. Kephart and David M. Chess. The Vision of Autonomic Computing. *Computer*, 36 (1):41–50, 2003.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations*, 12 2015. URL http://arxiv.org/abs/1412.6980.

Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 12 2014. URL http://arxiv.org/abs/1312.6114.

M. Klein-Wolt, R. P. Fender, G. G. Pooley, T. Belloni, S. Migliari, E. H. Morgan, and M. van der Klis. Hard X-ray states and radio emission in GRS 1915+105. *Monthly Notices of the Royal Astronomical Society*, 331(3):745–764, 4 2002. ISSN 00358711. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1046/j.1365-8711.2002.05223.x.

E. Kuulkers, A. Lutovinov, A. Parmar, F. Capitanio, N. Mowlavi, and W. Hermsen. Igr J17091-3624. *The Astronomer's Telegram*, 149:1, 2003.

Martin Längkvist, Lars Karlsson, and Amy Loutfi. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42(1):11–24, 6 2014. ISSN 01678655. .

Alan M Levine, Hale Bradt, Wei Cui, J G Jernigan, Edward H Morgan, Ronald Remillard, Robert E Shirey, and Donald A Smith. FIRST RESULTS FROM THE ALL-SKY MONITOR

ON THE ROSSI X-RAY TIMING EXPLORER. *THE ASTROPHYSICAL JOURNAL Letters*, 469:L33–L36, 1996.

Walter H. G. Lewin and Edward van den Heuvel. Accretion-driven stellar X-ray sources. 1983.

Walter H. G. Lewin, Jan van Paradijs, and Ronald E. Taam. X-Ray Bursts. *Space Science Reviews*, 63(1):223–389, 1993.

Jie Lin, Chengyuan Wu, Xiaofeng Wang, Péter Németh, Heran Xiong, Tao Wu, Alexei V Filippenko, Yongzhi Cai, Thomas G Brink, Shengyu Yan, Xiangyun Zeng, Yangping Luo, Danfeng Xiang, Jujia Zhang, WeiKang Zheng, Yi Yang, Jun Mo, Gaobo Xi, Jicheng Zhang, Abdusamatjan Iskandar, Ali Esamdin, Xiaojun Jiang, Hanna Sai, Zixuan Wei, Liyang Chen, Fangzhou Guo, Zhihao Chen, Wenxiong Li, Weili Lin, Han Lin, and Xinghan Zhang. An 18.9 min blue large-amplitude pulsator crossing the 'Hertzsprung gap' of hot subdwarfs. *Nature Astronomy*, 7:223–233, 2 2023. .

Yu Wei Lin, Yuqian Zhou, Faraz Faghri, Michael J. Shaw, and Roy H. Campbell. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long shortterm memory. *PLoS ONE*, 14(7), 7 2019. ISSN 19326203. .

N R Lomb. LEAST-SQUARES FREQUENCY ANALYSIS OF UNEQUALLY SPACED DATA. *Astrophys Space Sci*, 39(2):447–462, 1976.

Cristóbal Mackenzie, Karim Pichara, and Pavlos Protopapas. CLUSTERING-BASED FEATURE LEARNING ON VARIABLE STARS. *The Astrophysical Journal*, 820(2):138, 3 2016. ISSN 0004-637X. .

Ashish Mahabal, Kshiteej Sheth, Fabian Gieseke, Akshay Pai, S. George Djorgovski, Andrew Drake, Matthew Graham, and the CSS/CRTS/PTF Collaboration. Deep-Learnt Classification of Light Curves. In *IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 2757–2764, 9 2017. .

J. A. Martín-Fernández, C. Barceló-Vidal, and V. Pawlowsky-Glahn. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Mathematical Geology*, 35(3):253–278, 4 2003. ISSN 08828121. .

E. Massaro, F. Capitanio, M. Feroci, T. Mineo, A. Ardito, and P. Ricciardi. A non-linear mathematical model for the X-ray variability classes of the microquasar GRS 1915+105 – II: transition and swaying classes. *Monthly Notices of the Royal Astronomical Society*, 496(2): 1697–1705, 9 2020. . URL http://arxiv.org/abs/2009.07678http://dx.doi.org/10.1093/mnras/staa1125.

Masaru Matsuoka, Kazuyoshi Kawasaki, Shiro Ueno, Hiroshi Tomida, Mitsuhiro Kohama, Motoko Suzuki, Yasuki Adachi, Masaki Ishikawa, Tatehiro Mihara, Mutsumi Sugizaki, Naoki Isobe, Yujin Nakagawa, Hiroshi Tsunemi, Emi Miyata, Nobuyuki Kawai, Jun Kataoka, Mikio Morii, Atsumasa Yoshida, Hitoshi Negoro, Motoki Nakajima, Yoshihiro Ueda, Hirotaka

Chujo, Kazutaka Yamaoka, Osamu Yamazaki, Satoshi Nakahira, Tetsuya You, Ryoji Ishiwata, Sho Miyoshi, Satoshi Eguchi, Kazuo Hiroi, Haruyoshi Katayama, and Ken Ebisawa. The MAXI Mission on the ISS: Science and Instruments for Monitoring All-Sky X-Ray Images. *Publications of the Astronomical Society of Japan*, 61:999–1010, 2009. URL https://academic.oup.com/pasj/article/61/5/999/1593752.

Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

K. E. McGowan, M. J. Coe, M. P.E. Schurch, R. H.D. Corbet, J. L. Galache, and A. Udalski. Optical and X-ray variability of two Small Magellanic Cloud X-ray binary pulsars - SXP46.6 and SXP6.85. *Monthly Notices of the Royal Astronomical Society*, 384(2):821–826, 2008. ISSN 13652966. .

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, 2 2018. URL http://arxiv.org/abs/1802.03426.

J. M. Miller and J Homan. The Black Hole GRS 1915+105 is Still Active in X-rays. *The Astronomer's Telegram*, 15908:1, 2 2023.

I F Mirabel and L F Rodriguez. A superluminal source in the Galaxy. *Nature*, 371:46–48, 1994. .

S. E. Motta. Quasi periodic oscillations in black hole binaries. *Astronomische Nachrichten*, 337 (4-5):398–405, 3 2016. . URL http://arxiv.org/abs/1603.07885http://dx.doi.org/10.1002/asna.201612320.

Ashleigh Myall, James R. Price, Robert L. Peach, Mohamed Abbas, Sid Mookerjee, Nina Zhu, Isa Ahmad, Damien Ming, Farzan Ramzan, Daniel Teixeira, Christophe Graf, Andrea Y. Weiße, Stephan Harbarth, Alison Holmes, and Mauricio Barahona. Prediction of hospital-onset COVID-19 infections using dynamic networks of patient contact: an international retrospective cohort study. *The Lancet Digital Health*, 4(8):e573–e583, 8 2022. ISSN 25897500. .

S Naik, P P C Agrawal, P A R Rao, and B Paul. X-ray properties of the microquasar GRS 19151105 during a variability class transition. *Monthly Notices of the Royal Astronomical Society*, 330:487–496, 2002. . URL https://academic.oup.com/mnras/article/330/2/487/1085439.

Brett Naul, Joshua S. Bloom, Fernando Pérez, and Stéfan Van Der Walt. A recurrent neural network for classification of unevenly sampled variable stars. *Nature Astronomy*, 2(2):151–155, 2018. ISSN 23973366. .

Sergei Nayakshin, Saul Rappaport, and Fulvio Melia. Time-dependent Disk Models for the Microquasar GRS 1915+105. *The Astrophysical Journal*, 535(2):798–814, 6 2000. ISSN 0004-637X. . URL https://iopscience.iop.org/article/10.1086/308860.

Joseph Neilsen, Ronald A. Remillard, and Julia C. Lee. The physics of the "heartbeat" state of GRS1915+105. *Astrophysical Journal*, 737(2), 8 2011. ISSN 15384357. .

Isadora Nun, Pavlos Protopapas, Brandon Sim, Ming Zhu, Rahul Dave, Nicolas Castro, and Karim Pichara. FATS: Feature Analysis for Time Series, 5 2015. URL http://arxiv.org/abs/1506.00010.

Ruth Nussinov, Hyunbum Jang, Guy Nir, Chung Jung Tsai, and Feixiong Cheng. A new precision medicine initiative at the dawn of exascale computing. *Signal Transduction and Targeted Therapy*, 6(1), 12 2021. ISSN 20593635. .

Jakub K. Orwat-Kapola, Antony J. Bird, Adam B. Hill, Diego Altamirano, and Daniela Huppenkothen. Light-curve fingerprints: An automated approach to the extraction of X-ray variability patterns with feature aggregation - An example application to GRS 1915+105. *Monthly Notices of the Royal Astronomical Society*, 509(1):1269–1290, 1 2022. ISSN 13652966. .

B Paczynski. Gravitational Microlensing of the Galactic Bulge Stars. *Astrophysical Journal Letters*, 371:L63, 4 1991. .

Mayukh Pahari and Sabyasachi Pal. Signature of long-term class evolution in GRS1915+105 at a high accretion rate. *Monthly Notices of the Royal Astronomical Society*, 409(3):903–912, 12 2010. ISSN 00358711. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2010.17378.x.

P M Saz Parkinson, D M Tournear, E D Bloom, W B Focke, K T Reilly, K S Wood, P S Ray, M T Wolff, and Jeffrey D Scargle. LONG-TERM X-RAY VARIABILITY OF CIRCINUS X-1. *The Astrophysical Journal*, 595(1):333–341, 2003.

R. Pattnaik, K. Sharma, K. Alabarta, D. Altamirano, M. Chakraborty, A. Kembhavi, M. Méndez, and J. K. Orwat-Kapola. A machine-learning approach for classifying low-mass X-ray binaries based on their compact object nature. *Monthly Notices of the Royal Astronomical Society*, 501 (3):3457–3471, 3 2021. ISSN 13652966. .

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12: 2825–2830, 2011.

Christian Pieringer, Karim Pichara, Márcio Catelán, and Pavlos Protopapas. An Algorithm for the Visualization of Relevant Patterns in Astronomical Light Curves. *Monthly Notices of the Royal Astronomical Society*, 484(3):3071–3077, 3 2019. . URL http://arxiv.org/abs/1903.03254http://dx.doi.org/10.1093/mnras/stz106.

P. Praveen, C. H. Jayanth Babu, and B. Rama. Big data environment for geospatial data analysis. In *Proceedings of the International Conference on Communication and Electronics Systems, ICCES 2016*. Institute of Electrical and Electronics Engineers Inc., 2016. ISBN 9781509010653. .

M. Pursiainen, C. P. Gutiérrez, P. Wiseman, M. Childress, M. Smith, C. Frohmaier, C. Angus, N. Castro Segura, L. Kelsey, M. Sullivan, L. Galbany, P. Nugent, B. A. Bassett, D. Brout, D. Carollo, C. B. D'andrea, T. M. Davis, R. J. Foley, M. Grayling, S. R. Hinton, C. Inserra, R. Kessler, G. F. Lewis, C. Lidman, E. Macaulay, M. March, A. Möller, T. Müller, D. Scolnic, N. E. Sommer, E. Swann, B. P. Thomas, B. E. Tucker, M. Vincenzi, T. M.C. Abbott, S. Allam, J. Annis, S. Avila, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, L. N. Da Costa, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, T. F. Eifler, S. Everett, B. Flaugher, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. L. Hollowood, K. Honscheid, D. J. James, A. G. Kim, E. Krause, K. Kuehn, M. A.G. Maia, J. L. Marshall, F. Menanteau, R. Miquel, R. L.C. Ogando, A. Palmese, F. Paz-Chinchón, A. A. Plazas, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, E. Suchyta, M. E.C. Swanson, G. Tarle, and W. Wester. The mystery of photometric twins DES17X1boj and DES16E2bjy. *Monthly Notices of the Royal Astronomical Society*, 494(4):5576–5589, 6 2020. ISSN 13652966. .

N V Raguzova and S B Popov. Be/X-ray binaries and candidates. *Astron.Astrophys.Trans.*, 24 (3):151–185, 2005.

A. F. Rajoelimanana, P. A. Charles, and A. Udalski. Very long-term optical variability of high-mass X-ray binaries in the Small Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society*, 413(3):1600–1622, 5 2011. ISSN 00358711. .

M. J. Reid, J. E. McClintock, J. F. Steiner, D. Steeghs, R. A. Remillard, V. Dhawan, and R. Narayan. A parallax distance to the microquasar GRS 1915+105 and a revised estimate of its black hole mass. *Astrophysical Journal*, 796(1), 11 2014. ISSN 15384357. .

Pablo Reig. Be/X-ray binaries. *Astrophysics and Space Science*, 332(1):1–29, 2011. ISSN 0004640X. .

Ronald A. Remillard and Jeffrey E. McClintock. X-Ray Properties of Black-Hole Binaries. *Annual Review of Astronomy and Astrophysics*, 44(1):49–92, 2006. ISSN 0066-4146. .

Douglas A Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.

Joseph W. Richards, Dan L. Starr, Nathaniel R. Butler, Joshua S. Bloom, John M. Brewer, Arien Crellin-Quick, Justin Higgins, Rachel Kennedy, and Maxime Rischard. On Machine-Learned Classification of Variable Stars with Sparse and Noisy Time-Series Data. *The Astrophysical Journal*, 733(1), 1 2011. . URL http://dx.doi.org/10.1088/0004-637X/733/1/10.

Benjamin J. Ricketts, James F. Steiner, Cecilia Garraffo, Ronald A. Remillard, and Daniela Huppenkothen. Mapping the X-ray variability of GRS 1915 + 105 with machine learning. *Monthly Notices of the Royal Astronomical Society*, 523(2):1946–1966, 8 2023. ISSN 13652966. .

Thomas Rivinius, Alex C. Carciofi, and Christophe Martayan. Classical Be stars: Rapidly rotating B stars with viscous Keplerian decretion disks. *Astronomy and Astrophysics Review*, 21(1):1–86, 11 2013. ISSN 09354956. .

J Rodriguez, S Corbel, J. A. Tomsick, A Paizis, and E Kuulkers. 0.1 Hz QPOs during RXTE observations of IGR J17091-3624. *The Astronomer's Telegram*, 3168:1, 2 2011.

Lior Rokach and Oded Maimon. Clustering methods. In *The Data Mining and Knowledge Discovery Handbook*, pages 321–352. Springer, 2005.

Victor Rotaru, Yi Huang, Timmy Li, James Evans, Ishanu Chattopadhyay, and Health Sciences. Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. *Nature Human Behaviour*, 6(1):1056–1068, 2022.

Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*, 2017.

P. Sánchez-Sáez, H. Lira, L. Martí, N. Sánchez-Pi, J. Arredondo, F. E. Bauer, A. Bayo, G. Cabrera-Vives, C. Donoso-Oliva, P. A. Estévez, S. Eyheramendy, F. Förster, L. Hernández-García, A. M. Muñoz Arancibia, M. Pérez-Carrasco, M. Sepúlveda, and J. R. Vergara. Searching for Changing-state AGNs in Massive Data Sets. I. Applying Deep Learning and Anomaly-detection Techniques to Find AGNs with Anomalous Variability Behaviors. *The Astronomical Journal*, 162(5):206, 11 2021. ISSN 0004-6256. .

Jeffrey D Scargle. STUDIES IN ASTRONOMICAL TIME SERIES ANALYSIS. II. STATISTICAL ASPECTS OF SPECTRAL ANALYSIS OF UNEVENLY SPACED DATA. *The Astrophysical Journal*, 263:835–853, 1982.

S. Scaringi, D. de Martino, D. H. Buckley, P. J. Groot, C. Knigge, M. Fratta, K. Ilkiewicz, C. Littlefield, and A. Papitto. An accreting white dwarf displaying fast transitional mode switching. *Nature Astronomy*, 6(1):98–102, 10 2022. . URL http://dx.doi.org/10.1038/s41550-021-01494-x.

Ronald W. Schafer. What is a savitzky-golay filter? *IEEE Signal Processing Magazine*, 28(4): 111–117, 2011. ISSN 10535888. .

P. C. Schmidtke and A. P. Cowley. Photometric Periodicities of Be/X-Ray Pulsars in the Small Magellanic Cloud. *The Astronomical Journal*, 130(5):2220–2229, 11 2005. .

P. C. Schmidtke and A. P. Cowley. Photometric Periods for SMC Be/X-ray Pulsars SXP645 and SXP726. *The Astronomer's Telegram*, 1632:1, 7 2008.

P C Schmidtke, A P Cowley, Lance Levenson, and Katie Sweet. MULTIPERIODIC VARIABILITY IN MAGELLANIC CLOUD Be/X-RAY BINARIES 1. *The Astronomical Journal*, 127 (6):3388–3393, 2004.

P C Schmidtke, A P Cowley, and A Udalski. AN INVESTIGATION OF Be/X-RAY PULSARS WITH OGLE-III DATA. *The Astronomical Journal*, 132(3):971, 2006. URL http://www.astro.soton.ac.uk/~mjc/smc.

P. C. Schmidtke, A. P. Cowley, and A Udalski. Photometric Period of XTE J0053-724 (SXP46.6 = 1WGA J0053.8-7226). *The Astronomer's Telegram*, 1316:1, 11 2007.

P. C. Schmidtke, A. P. Cowley, and A. Udalski. An analysis of optical periods in individual small magellanic cloud be/x-ray pulsars. *Monthly Notices of the Royal Astronomical Society*, 431 (1):252–278, 5 2013. ISSN 13652966. .

Neil Shah, Nandish Bhagat, and Manan Shah. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. *Visual Computing for Industry, Biomedicine, and Art*, 4(1):1–14, 12 2021. ISSN 25244442. .

Christopher J. Shallue and Andrew Vanderburg. Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *The Astronomical Journal*, 155(2):94, 1 2018. ISSN 1538-3881. .

Shailendra Singh and Abdulsalam Yassine. Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, 11(2), 2 2018. ISSN 19961073. .

T M Tauris and E P J Van Den Heuvel. Formation and Evolution of Compact Stellar X-ray Sources. In Michiel van der Klis and Walter Lewin, editors, *Compact stellar X-ray sources*, pages 623–665. 2003.

Sean J. Taylor and Benjamin Letham. Forecasting at Scale. *American Statistician*, 72(1):37–45, 1 2018. ISSN 15372731. .

Matthew Templeton. Time-Series Analysis of Variable Star Data. *The Journal of the American Association of Variable Star Observers*, 32(1):41–54, 6 2004.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000. . URL https://www.science.org/doi/abs/10.1126/science.290.5500.2319.

The Astropy Collaboration, Adrian M. Price-Whelan, Pey Lian Lim, Nicholas Earl, Nathaniel Starkman, Larry Bradley, David L. Shupe, Aarya A. Patil, Lia Corrales, C. E. Brasseur, Maximilian Nöthe, Axel Donath, Erik Tollerud, Brett M. Morris, Adam Ginsburg, Eero Vaher, Benjamin A. Weaver, James Tocknell, William Jamieson, Marten H. van Kerkwijk, Thomas P. Robitaille, Bruce Merry, Matteo Bachetti, H. Moritz Günther, Thomas L. Aldcroft, Jaime A. Alvarado-Montes, Anne M. Archibald, Attila Bódi, Shreyas Bapat, Geert Barentsen, Juanjo Bazán, Manish Biswas, Médéric Boquien, D. J. Burke, Daria Cara, Mihai Cara, Kyle E Conroy, Simon Conseil, Matthew W. Craig, Robert M. Cross, Kelle L. Cruz, Francesco D'Eugenio, Nadia Dencheva, Hadrien A. R. Devillepoix, Jörg P. Dietrich, Arthur Davis Eigenbrot, Thomas Erben, Leonardo Ferreira, Daniel Foreman-Mackey, Ryan Fox, Nabil Freij, Suyog Garg, Robel Geda, Lauren Glattly, Yash Gondhalekar, Karl D. Gordon, David Grant, Perry Greenfield, Austen M. Groener, Steve Guest, Sebastian Gurovich, Rasmus Handberg, Akeem Hart, Zac Hatfield-Dodds, Derek Homeier, Griffin Hosseinzadeh, Tim Jenness, Craig K. Jones, Prajwel Joseph, J. Bryce Kalmbach, Emir Karamehmetoglu, Mikołaj Kałuszyński, Michael S. P. Kelley, Nicholas Kern, Wolfgang E. Kerzendorf, Eric W. Koch, Shankar Kulumani, Antony Lee, Chun Ly, Zhiyuan Ma, Conor MacBride, Jakob M.

Maljaars, Demitri Muna, N. A. Murphy, Henrik Norman, Richard O'Steen, Kyle A. Oman, Camilla Pacifici, Sergio Pascual, J. Pascual-Granado, Rohit R. Patil, Gabriel I Perren, Timothy E. Pickering, Tanuj Rastogi, Benjamin R. Roulston, Daniel F Ryan, Eli S. Rykoff, Jose Sabater, Parikshit Sakurikar, Jesús Salgado, Aniket Sanghi, Nicholas Saunders, Volodymyr Savchenko, Ludwig Schwardt, Michael Seifert-Eckert, Albert Y. Shih, Anany Shrey Jain, Gyanendra Shukla, Jonathan Sick, Chris Simpson, Sudheesh Singanamalla, Leo P. Singer, Jaladh Singhal, Manodeep Sinha, Brigitta M. Sipőcz, Lee R. Spitler, David Stansby, Ole Streicher, Jani Šumak, John D. Swinbank, Dan S. Taranu, Nikita Tewary, Grant R. Tremblay, Miguel de Val-Borro, Samuel J. Van Kooten, Zlatan Vasović, Shresth Verma, José Vinícius de Miranda Cardoso, Peter K. G. Williams, Tom J. Wilson, Benjamin Winkel, W. M. Wood-Vasey, Rui Xue, Peter Yoachim, Chen ZHANG, and Andrea Zonca. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *The Astrophysical Journal*, 935(167), 6 2022. . URL http://arxiv.org/abs/2206.14220http://dx.doi.org/10.3847/1538-4357/ac7c74.

L. J. Townsend, S. P. Drave, A. B. Hill, M. J. Coe, R. H.D. Corbet, A. J. Bird, and M. P.E. Schurch. Contrasting behaviour from two Be/X-ray binary pulsars: Insights into differing neutron star accretion modes. *Monthly Notices of the Royal Astronomical Society*, 433(1): 23–35, 2013. ISSN 13652966. .

A Udalski, M Szymanski, J Kaluzny, M Kubiak, and Mario Mateo. The Optical Gravitational Lensing Experiment. *Acta Astronomica*, 42:253–284, 10 1992.

A. Udalski, M. Kubiak, and M. Szymanski. Optical Gravitational Lensing Experiment. OGLE-2-the Second Phase of the OGLE Project. *Acta Astronomica*, 47:319–344, 1997.

A. Udalski, M. K. Szymański, and G. Szymański. OGLE-IV: Fourth Phase of the Optical Gravitational Lensing Experiment. *Acta Astronomica*, 65:1–38, 4 2015. URL http://arxiv.org/abs/1504.05966.

Andrzej Udalski. The Optical Gravitational Lensing Experiment. Real Time Data Analysis Systems in the OGLE-III Survey 1. *Acta Astronomica*, 53:291–305, 2003.

Andrzej Udalski. The Optical Gravitational Lensing Experiment . OGLE-III Photometric Maps of the Large Magellanic Cloud. *Acta Astronomica*, 58:329–343, 2008.

Ryan T. Urquhart, Roberto Soria, Rosanne Di Stefano, Kaiming Cui, Paolo Esposito, Gian Luca Israel, Sammarth Kumar, Sara Motta, Fabio Pintore, and Giacomo Riva. Quasi periodic whispers from a transient ULX in M\,101: signatures of a fast-spinning neutron star? *Monthly Notices of the Royal Astronomical Society*, 511(3):4528–4550, 1 2022. URL http://arxiv.org/abs/2201.07252.

Lucas Valenzuela and Karim Pichara. Unsupervised Classification of Variable Stars. *Monthly Notices of the Royal Astronomical Society*, 474(3):3259–3272, 1 2018. . URL http://arxiv.org/abs/1801.09723http://dx.doi.org/10.1093/mnras/stx2913.

Laurens Van Der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

Jacob Vanderplas, Andrew J. Connolly, Zeljko Ivezic, and Alex Gray. Introduction to astroML: Machine learning for astrophysics. In *Proceedings - 2012 Conference on Intelligent Data Understanding, CIDU 2012*, pages 47–54, 2012. ISBN 9781467346252. .

Jacob T. VanderPlas. Understanding the Lomb–Scargle Periodogram. *The Astrophysical Journal Supplement Series*, 236(1):16, 5 2018. ISSN 00670049. .

Jacob T. Vanderplas and Željko Ivezić. PERIODOGRAMS for MULTIBAND ASTRONOMICAL TIME SERIES. *Astrophysical Journal*, 812(1), 10 2015. ISSN 15384357. .

Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, and Andras Anderla. Credit Card Fraud Detection - Machine Learning methods. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–5, 2019. .

S. Vaughan, R. Edelson, R. S. Warwick, and P. Uttley. On characterizing the variability properties of X-ray light curves from active galaxies. *Monthly Notices of the Royal Astronomical Society*, 345(4):1271–1284, 2003. ISSN 00358711. .

Frank Verbunt. ORIGIN AND EVOLUTION OF X-RAY BINARIES AND BINARY RADIO PULSARS. *Annu. Rev. Astron. Astrophys.*, 31:93–127, 1993. URL www.annualreviews.org.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R.J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T. Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones,

Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 3 2020. ISSN 15487105. .

Paul J. Werbos. Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE*, 78(10):1550–1560, 1990. ISSN 15582256. .

John Whelan. BINARIES AND SUPERNOVAE OF TYPE I*. *The Astrophysical Journal*, 186: 1007–1014, 1973.

P. Wiseman, M. Pursiainen, M. Childress, E. Swann, M. Smith, L. Galbany, C. Lidman, T. M. Davis, C. P. Gutiérrez, A. Möller, B. P. Thomas, C. Frohmaier, R. J. Foley, S. R. Hinton, L. Kelsey, R. Kessler, G. F. Lewis, M. Sako, D. Scolnic, M. Sullivan, M. Vincenzi, T. M.C. Abbott, M. Aguena, S. Allam, J. Annis, E. Bertin, S. Bhargava, D. Brooks, D. L. Burke, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, J. Carretero, M. Costanzi, L. N. da Costa, H. T. Diehl, P. Doel, S. Everett, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, K. Glazebrook, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, M. Lima, M. A.G. Maia, J. L. Marshall, P. Martini, F. Menanteau, R. Miquel, A. Palmese, F. Paz-Chinchón, A. A. Plazas, A. K. Romer, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, N. E. Sommer, E. Suchyta, M. E.C. Swanson, G. Tarle, B. E. Tucker, D. L. Tucker, T. N. Varga, and A. R. Walker. The host galaxies of 106 rapidly evolving transients discovered by the Dark Energy Survey. *Monthly Notices of the Royal Astronomical Society*, 498(2):2575–2593, 2020. ISSN 13652966. .

Xuhua Xia. Bioinformatics and Drug Discovery. *Current Topics in Medicinal Chemistry*, 17 (15):1709–1726, 4 2017. ISSN 15680266. .

J. Yang, S. G. T. Laycock, D. M. Christodoulou, S. Fingerman, M. J. Coe, and J. J. Drake. A Comprehensive Library of X-Ray Pulsars in the Small Magellanic Cloud: Time Evolution of Their Luminosities and Spin Periods. *The Astrophysical Journal*, 839(2):119, 4 2017. ISSN 15384357. .

Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural Computation*, 31(7):1235–1270, 7 2019. ISSN 1530888X. .

M. Zechmeister and M. Kürster. The generalised Lomb-Scargle periodogram a new formalism for the floating-mean and Keplerian periodograms. *Astronomy and Astrophysics*, 496(2):577–584, 3 2009. ISSN 00046361. .

Keming Zhang and Joshua S. Bloom. Classification of periodic variable stars with novel cyclic-permutation invariant neural networks. *Monthly Notices of the Royal Astronomical Society*, 505(1):515–522, 7 2021. ISSN 13652966. .

Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.

Li Zhu, Fei Richard Yu, Yige Wang, Bin Ning, and Tao Tang. Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1):383–398, 1 2019. ISSN 15249050. .