

Shaplet method to classify GRS1915+105 black hole signal anomalies

Bouزيد Aymane

Supervised by Gangler emmanuel

Contents

1	Object of Study	3
1.1	Introduction to X binaries	3
1.2	Definitions:	3
1.2.1	The Physics Behind the Radiation	4
1.2.2	Detection Techniques	5
1.2.3	Introduction to GRS 1915+105	6
1.3	Previous Work	7
2	Shaplets in action	10
2.1	Shaplet Algorithm	10
2.1.1	Advantages of Shapelets:	10
2.2	Variants Used in Our Work	11
2.3	Shapelet Choice	11
2.4	Chi-Square Test and Empirical Validation	12
2.5	The Minimum Metric and Its Distribution	13
2.6	Finding a Cut and Efficient Cut Concept	14
2.7	Robustness Testing	15
2.8	Averaging	16
2.9	Scaling and normalisation	17
3	Towards a Dictionary of Shaplets	18
3.1	Shaplet Selection and Recursive Sampling Strategy	18

Acknowledgments

First and foremost, I would like to sincerely thank Prof. Emmanuel Gangler, who made this internship possible. His pedagogy, constant enthusiasm, and kindness in answering questions—no matter how frequent—made this experience deeply enriching. His patience and encouragement were invaluable, even during moments when I was not at my best.

I am also very grateful to the entire LPCA-Fink team, especially Dr. Emille Eugenia de Oliveira Ishida, the team leader, whose warm welcome and supportive presence were deeply appreciated throughout my stay.

A special thank you to my fellow interns at Fink, particularly Clément Mur, my officemate, and Eva de Araujo Pereira, for their camaraderie and collaboration, which made the work environment both productive and enjoyable.

Chapter 1

Object of Study

1.1 Introduction to X binaries

An X-ray binary is a system composed of a non-degenerate star and a compact object (either a black hole, a neutron star, or, rarely, a white dwarf) that accretes matter from the star via Roche lobe overflow or stellar wind [4]. The accretion process produces strong X-ray emissions, typically in the range of 10^{36} to 10^{38} erg/s, making these among the brightest stellar-mass X-ray sources. Key observables include the orbital period (ranging from a few hours to several days), the X-ray to optical flux ratio (often > 1000 in low-mass systems), and the light curve, which reveals variability patterns such as dips, eclipses, or bursts. These observables enable estimates of the compact object's mass (e.g., $> 3 M_{\odot}$ for black holes, $1.4\text{--}2.5 M_{\odot}$ for neutron stars) and the companion star's mass. X-ray binaries can be classified as:

- **Microquasars:** X-ray binaries with relativistic jets ($v \sim 0.9c$), typically involving a black hole or neutron star with an accretion disk.
- **Pulsar X-ray Binaries:** Systems with a rotating, magnetized neutron star producing coherent X-ray pulsations.
- **Transient X-ray Binaries:** Show occasional outbursts with luminosity rising from $L_X \sim 10^{33}$ to 10^{38} erg/s, due to disk instabilities.
- **Persistent X-ray Binaries:** Emit steady X-rays ($L_X \gtrsim 10^{36}$ erg/s) from continuous accretion.
- **Ultra-compact X-ray Binaries (UCXBs):** Have orbital periods < 80 minutes, with compact, hydrogen-poor donors like white dwarfs.

1.2 Definitions:

Light curve: shows how brightness changes over time revealing patterns like dips, bursts, or eclipses

Color: compares brightness in two filters or energy bands to indicate temperature and material type

Flux: measures the amount of energy received from the source, with higher flux meaning brighter emission

Magnitude: a scale for brightness where smaller numbers indicate brighter objects

Color index: the difference between magnitudes in two bands, related to the temperature and energy distribution

1.2.1 The Physics Behind the Radiation

Key Equations and Principles of Accretion Physics

We consider a Keplerian accretion disk of gas orbiting a central mass M , assuming:

- The self-gravity of the disk is negligible compared to the central object.
- The disk is geometrically thin and relatively cool.

The specific mechanical energy E of a particle of mass m at radius r is:

$$E = -\frac{GMm}{2r}$$

In the presence of viscous dissipation, angular momentum is transported outward, and the particle spirals inward, converting gravitational potential energy into radiation.

The total energy released per unit mass as matter falls from an external radius r_e to an inner radius r_i is:

$$\Delta E = \frac{GM}{2r_i} \left(1 - \frac{r_i}{r_e} \right)$$

If \dot{M} is the mass accretion rate, then the total power radiated (luminosity) by the disk is:

$$L = \Delta E \cdot \dot{M} = \frac{GM\dot{M}}{2r_i}$$

For a Schwarzschild black hole, the innermost stable circular orbit (ISCO) is at $r_i = 3R_S = \frac{6GM}{c^2}$. In that case, the radiated luminosity becomes:

$$L = \frac{1}{12} \dot{M} c^2$$

This shows the remarkable efficiency of accretion, converting up to $\sim 8\%$ of rest mass energy into radiation—orders of magnitude higher than nuclear fusion.

Disk Temperature Profile

Assuming the disk is optically thick and radiates as a blackbody, the effective temperature $T_{\text{eff}}(r)$ of a ring at radius r is given by:

$$T_{\text{eff}}(r) \propto \left(\frac{GM\dot{M}}{r^3} \right)^{1/4}$$

This implies:

- Inner regions are hotter and radiate in X-rays (for stellar-mass black holes).
- Outer regions emit in visible or UV (for supermassive black holes).

1.2.2 Detection Techniques

Radiation from X-ray binaries—especially microquasars—can span a broad range of the electromagnetic spectrum. However, this report focuses primarily on X-ray detection, as the light curves analyzed are derived from X-ray observations (the definition and properties of light curves are discussed in a previous section).

X-rays emitted from X-ray binaries typically range from a few keV to around 200 keV. Within this energy domain, the dominant photon-matter interaction mechanisms include the **photoelectric effect** and **Compton scattering**, with the photoelectric effect prevailing at lower energies (up to ~ 100 keV). **Rayleigh scattering** may also occur, though it is rare and generally negligible for detection purposes.

RXTE Detectors

According to documentation on NASA’s RXTE website, the Rossi X-ray Timing Explorer (RXTE) was equipped with two primary X-ray instruments: the **Proportional Counter Array (PCA)** and the **High Energy X-ray Timing Experiment (HEXTE)**. The PCA included five xenon-filled Proportional Counter Units (PCUs) covering the 2–60 keV range. Each PCU featured a multi-layer design with xenon gas detection volumes, entrance windows, and anode wires. **Veto layers** filled with propane gas were incorporated to reduce background noise.

In both PCA and HEXTE, the dominant detection mechanism is the **photoelectric effect**. In PCA, incident X-ray photons ionize xenon atoms, and the freed electrons are collected at the anode wires, generating a signal proportional to the photon’s energy. In HEXTE, the photoelectric effect also dominates at lower energies, while at energies above ~ 100 keV, **Compton scattering** contributes to the energy deposition.

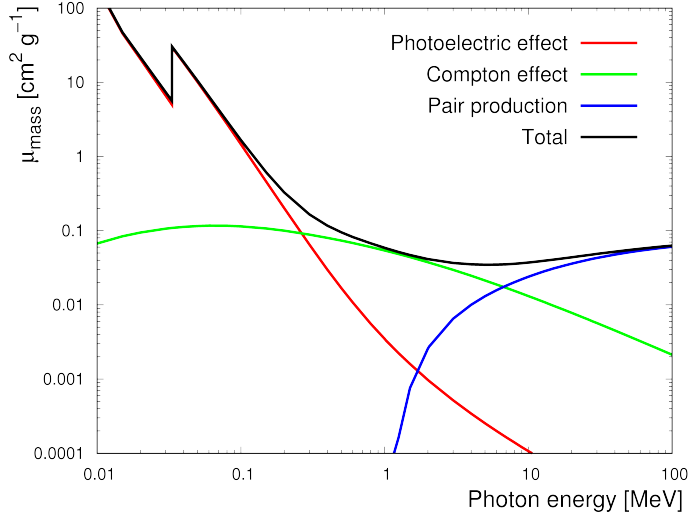


Figure 1.1: Plot illustrating the dominant photon and/or X-ray interaction mechanisms as a function of energy. Source: INFN Ferrara.

Data Processing and Analysis

Information from PCA detectors is handled by the **Experiment Data System (EDS)**, an onboard microprocessor capable of processing up to 500,000 X-ray events per second with a timing precision of 1 microsecond. The EDS supports multiple data collection modes simultaneously, tailored to different observational goals. HEXTE uses a separate but functionally similar data system.

The collected data are stored in distinct files corresponding to different observation modes. From these, astronomers extract **light curves**, which represent intensity variations over time, and **energy spectra**, which show photon count as a function of energy. Light curves are analyzed to identify variability, periodic behavior, or complex modulations—offering insights into physical processes like accretion or orbital motion. Spectral data are fit using physical models to derive properties such as luminosity, energy emission mechanisms, and elemental composition.

The dataset used in this study is available at the same link cited in [3]. It contains all the light curves, each separated into three energy bands: low, medium, and high.

1.2.3 Introduction to GRS 1915+105

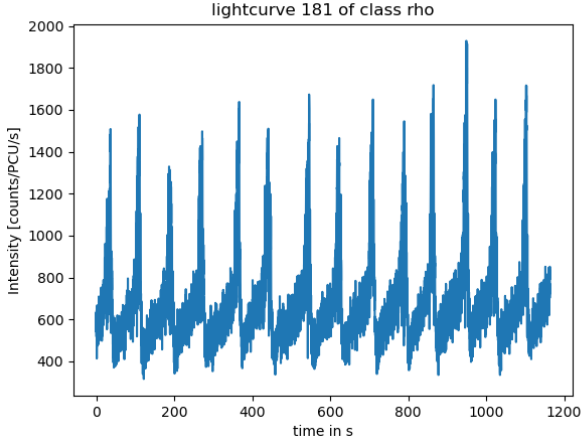
GRS 1915+105 is a microquasar [2] located in our galaxy, approximately 11,000 parsecs away from the Solar System [2], in the constellation Aquila [1]. It consists of a stellar black hole and a low-mass main sequence star [2]. The companion star orbits the black hole with a period of 33.5 days in a relatively wide orbit. The black hole accretes matter from its companion, forming an accretion disk. As the material spirals inward, it heats up and emits X-rays. GRS 1915+105 was discovered by the French-Russian space satellite Granat on August 15, 1992. The data analyzed in this report comes from RXTE. In 1994, GRS 1915+105 became the first known

galactic object to eject material at velocities that appear superluminal—faster than the speed of light—due to relativistic effects. The name "GRS" stands for "GRANAT source." The number "1915" refers to its right ascension 19 hours and 15 minutes, while "105" reflects its approximate declination 10 degrees and 56 arcminutes. What distinguishes GRS 1915+105 from other known black hole X-ray binaries (BHXRBS) is its extraordinary X-ray variability.[3] While variability in flux and spectrum is generally expected due to turbulence in the accretion disk—likely caused by magnetic instabilities—GRS 1915+105 exhibits a remarkable range of complexity. Its X-ray light curves fall into at least 14 distinct patterns, as identified by Belloni et al. (2000), Klein-Wolt et al. (2002), and Hannikainen et al. (2003, 2005). These patterns are not random; they tend to repeat almost identically, sometimes with months or even years between occurrences. Initially believed to be unique in this behavior, GRS 1915+105 found a counterpart in 2011 when IGR J170913624 was observed to exhibit similar variability (Altamirano et al., 2011). This variability is tightly linked with rapid spectral changes, posing challenges to standard accretion theories. Yet, decoding the origin of these patterns is essential, as they provide insight into the physical processes occurring in the accretion disk ref 1.3

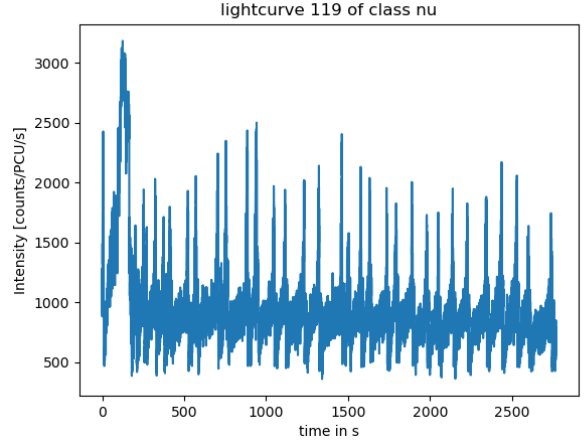
1.3 Previous Work

Outbursts in X-ray binaries are sudden increases in brightness due to rapid inflows of matter onto a compact object like a black hole. These events often mark transitions between different accretion states. Identifying and tracking outburst modes ¹ helps organize the complex behavior of such systems into understandable categories. Classifying outbursts and modes; especially using machine learning, helps researchers uncover how physical changes in the disk lead to different observable patterns. For example, a switch from one mode to another might indicate the onset of magnetic instabilities or a change in the inner disk radius. By grouping similar light curves and spectra across many observations, it is possible to isolate which physical conditions produce which behaviors. This makes it possible to test and refine theoretical models of accretion and jet formation. In essence, classification translates raw observational data into physical insight, providing a bridge between what we see and how these systems actually work. Using data from RXTE, astronomers observed that GRS 1915+105 exhibits uniquely complex and repeating X-ray light curves, grouped into at least 14 distinct patterns. Unlike typical black hole X-ray binaries, which usually transition between just three standard spectral states, GRS 1915+105 showed a much richer and more structured variability. For over a decade after its discovery in 1992, it was the only known microquasar to behave this way. It wasn't until 2006, with the discovery of similar variability in IGR J17091–3624, that a second source was found to exhibit comparable behavior—highlighting just how rare and exceptional these patterns are. The fact that GRS 1915+105's light curve patterns repeat almost identically over months or years suggests that they are governed by robust, underlying physical processes. Classifying

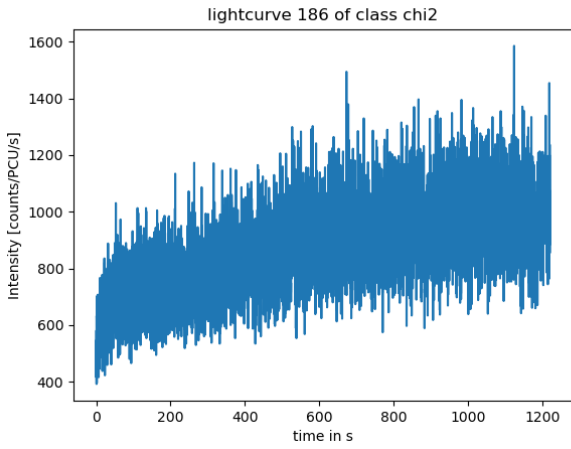
¹modes refer to distinct patterns of X-ray variability and spectral properties that reflect changes in the physical conditions of the accretion disk, such as its temperature, density, or level of turbulence



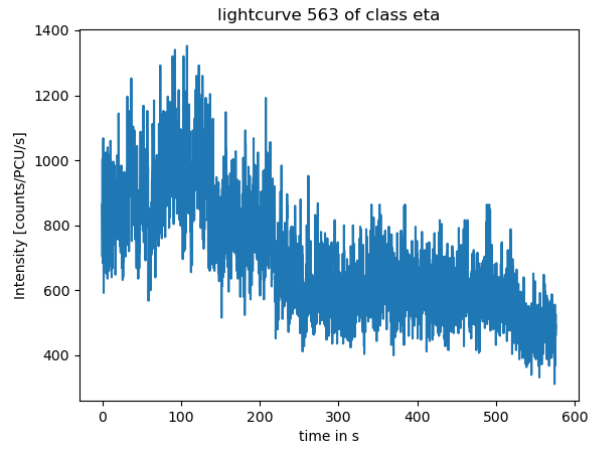
(a) lightcurve 181 of class rho



(b) lightcurve 119 of class nu



(c) lightcurve 186 of class chi2



(d) lightcurve 563 of class eta

Figure 1.2: Four examples of light curves from different classes. Light curve (a) represents class rho lightcurves, characterized by quasi-periodic behavior, making it an ideal test case for our algorithms.

these patterns is crucial for linking them to disk instabilities and accretion dynamics, making GRS 1915+105 a unique and invaluable case study in black hole astrophysics.

In the context of GRS 1915+105, researchers used machine learning tools such as autoencoders² to study its brightness and color-color variability, hoping to uncover hidden structure in its X-ray behavior. They compared the groupings identified by the autoencoder with those made by human experts, who typically classify light curves based on visual patterns and spectral characteristics. This side-by-side comparison allowed them to evaluate how well human-defined classes align with the machine’s learned features, and whether the complex classifications of GRS 1915+105 could be reduced to simpler, more objective categories.

However, a major limitation of neural networks, including autoencoders, is their lack of interpretability—they often act as “black boxes,” making it difficult to understand how or why specific decisions are made. In physics and astrophysics, where understanding the connection between observations and physical processes is essential, this lack of transparency poses a problem. Interpretable models are crucial for linking patterns to underlying physics, validating results, and ensuring scientific reliability. This is where our work comes in: we aim to develop a classification method that is both accurate and interpretable, allowing us to identify meaningful patterns in GRS 1915+105’s variability while retaining a clear connection to physical processes.

²A neural network autoencoder is a type of artificial neural network used to learn efficient representations of input data, typically for dimensionality reduction or denoising, by training the network to reconstruct the input from a compressed version of itself.

Chapter 2

Shaplets in action

2.1 Shaplet Algorithm

The shapelet algorithm is a supervised time series ¹ classification method introduced by Lexiang Ye and Eamonn Keogh [PDF]. A key advantage over traditional methods like nearest neighbor is its ability to yield more *interpretable results*. This is particularly valuable in computational physics, where the goal is not only accurate classification but also an understanding of the decision-making process.

In our case, we aim not only to improve upon the classification results of previous studies, but also to identify the hidden criteria that the algorithm relies on. These insights can potentially be connected to physical models describing phenomena such as outbursts.

We begin by defining shapelets. A shapelet is a subsequence of a time series that is maximally representative of a particular class. For instance, consider the classification of two morphologically similar plant species: *Urtica dioica* and *Verbena urticifolia*. The differences in their global shapes are subtle, and distortions such as insect damage may confuse traditional global classifiers.

To overcome this, the shapelet method first transforms each leaf’s outline into a one-dimensional time series. Instead of comparing entire shapes, it searches for short, local subsequences that are highly discriminative; these are the shapelets. In one example, the shapelet discovered reveals that *Urtica dioica* has a stem connecting to the leaf at nearly 90 degrees, whereas in *Verbena urticifolia* the connection is at a shallower angle.

2.1.1 Advantages of Shapelets:

- **Interpretability:** The algorithm extracts interpretable features, such as specific angles or curves, that provide insight into the classification rationale.
- **Robustness and Accuracy:** Because shapelets focus on local subsequences, they are more resistant to noise and global artifacts than global feature-based methods.

¹A time series is a sequence of data points indexed in time order, spaced at uniform intervals.

- **Computational Efficiency:** Classification has a time complexity of $\mathcal{O}(ml)$, where m is the length of the query time series and l is the length of the shapelet.
- **Time translation invariant:** This is because the shapelet is slid across the entire signal, and the best local match is used for comparison. As a result, the method is more flexible and accurate, especially when the key pattern doesn't always occur at the same time point

2.2 Variants Used in Our Work

We begin with preliminary tests using a single shapelet, mainly to evaluate whether it can produce meaningful and coherent results. To address this, we start by working with the existing classification, from which we randomly choose an initial shapelet. This initial choice serves as a starting point to iteratively construct a dictionary of shapelets that can represent all target classes. Further details regarding this approach are presented in Chapter 3. Despite adapting the methodology, we preserve the core principle of the shapelet framework: representing light curves using discriminative local patterns. In the following sections, we describe the comparison procedure and present preliminary results.

2.3 Shapelet Choice

Remark: Most of the analysis presented below is based on the ρ (rho) class. This choice is motivated by two main reasons: first, the ρ class exhibits quasi-periodic behavior, which makes it easier to manipulate and test the performance of our analysis methods; and second, the total number of light curves in this class is significantly smaller—approximately one-tenth—compared to the complete dataset, which substantially reduces computation time.

For the initial selection of shapelets, we used the two shapelets described below:

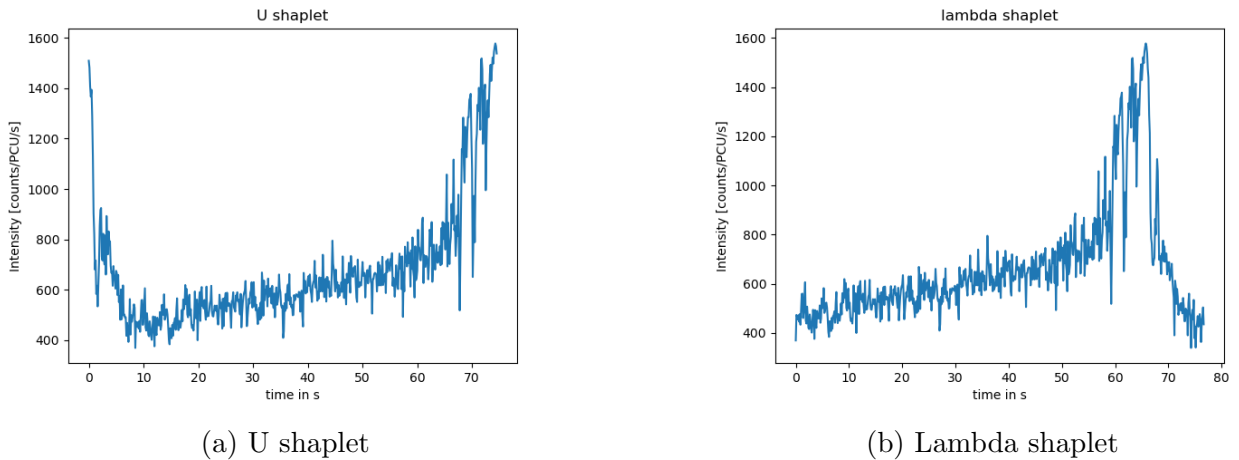


Figure 2.1: U shaplet and lambda shaplet both extracted from the lightcurve 181

We began with the U-shaped shapelet, as it is easier to define; the maxima are well-separated from the noise, while the minima tend to blend with background fluctuations, making them harder to isolate (i.e., fuzzy minima). These initial shapelets are 75 seconds long and contain 600 data points. Later in the study, we will transition to using shapelets of 32.5 seconds (300 points), as these shorter segments are more precise in capturing specific physical features, whereas the 600-point shapelets encapsulate more of the overall cycle characteristic of the periodic class.

2.4 Chi-Square Test and Empirical Validation

To compare a given shapelet with a light curve, we employ a *windowing* approach. This involves sliding the shapelet point-by-point along the light curve and recording a similarity score at each position. The similarity is quantified using a reduced chi-squared metric defined as:

$$\chi^2[d]/dof = \sum_{i=0}^N \frac{(L[i+d] - S[i])^2}{\sigma_S^2 + \sigma_L^2} / dof$$

where $L[i]$ and $S[i]$ are the light curve and shapelet values, respectively, and dof is the length of the shapelet.

A first simplification assumes $\sigma_L = \sigma_S$, justified by the null hypothesis that the shapelet and the matching segment of the light curve should be similar. Furthermore, considering the Poissonian nature of the signal, we approximate σ_S^2 as the mean of the shapelet.

With basic algebraic manipulation, possible because the denominator is constant, it can be shown that the chi-squared computation reduces to a convolution. This allows us to leverage the Fast Fourier Transform (FFT) for efficient computation.

After evaluating the results, we observe that the minimum value of the chi-squared curve is close to 1 for the majority of light curves, validating the consistency of the analysis.

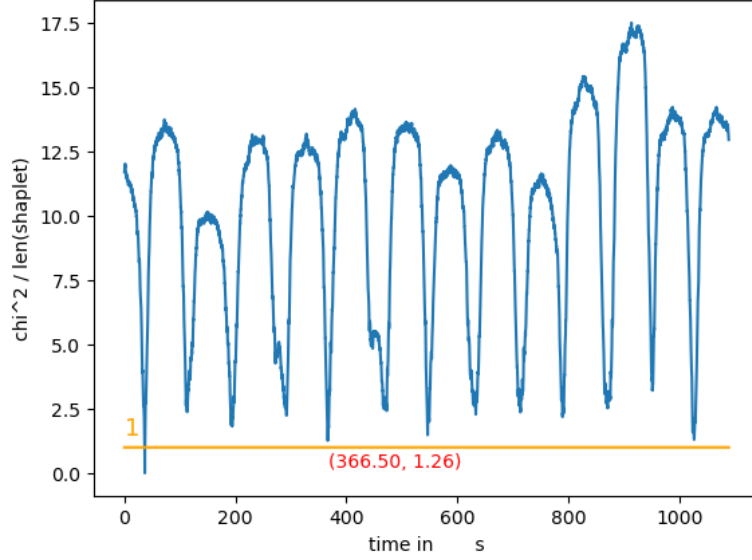


Figure 2.2: The plot shows the result of the windowing process. We observe that all minima are close to the orange line at value 1, which is expected from the χ^2 test. An exception occurs with LC181, where the minimum is 0. This is because the shaplet was extracted from LC181 itself, so windowing it at its original position yields a perfect match with zero distance. It's important to note that to get the minimas we use a detection window of half the size of the shaplet which is one of the free parameters to study further later on

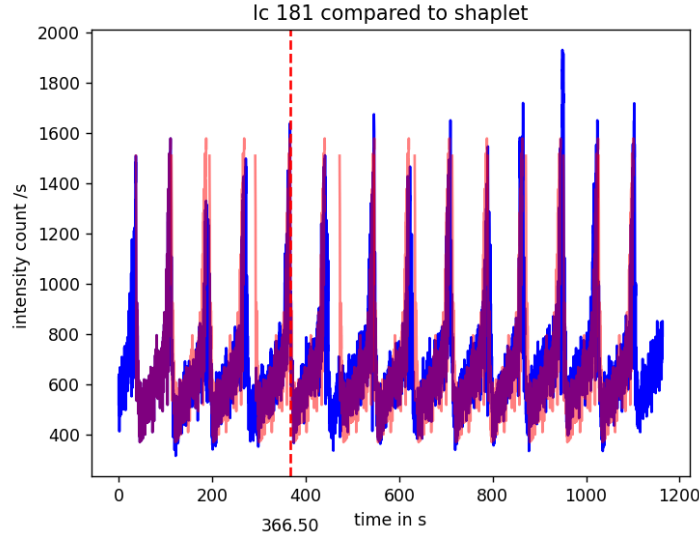
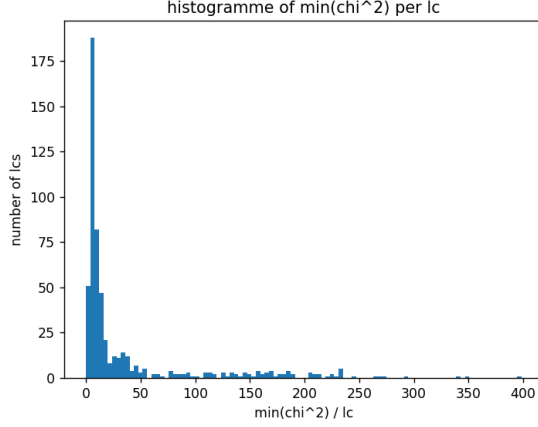


Figure 2.3: The result of superposing the shapelet on the light curve only at indices corresponding to minima shows close matches and a perfect match at 36.75 seconds, which is the starting point of the shapelet.

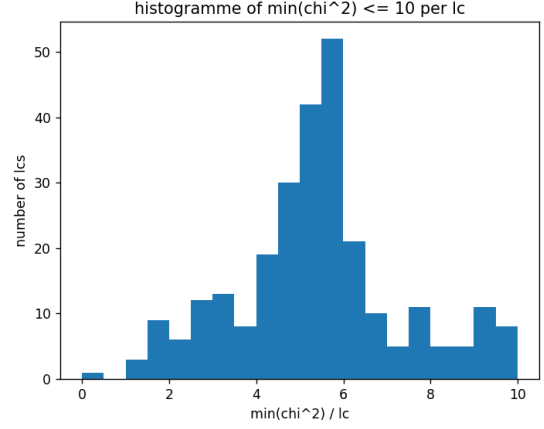
2.5 The Minimum Metric and Its Distribution

Since the core idea of the algorithm is to classify based on a single, class-defining pattern represented by the shapelet, it is essential to capture the **best local match** within the time series. The minimum χ^2 value obtained through windowing thus serves as a justified and

effective metric for quantifying the similarity between a shapelet and a signal. Below is the distribution obtained using this metric:



(a) Histogram of minimum values across all samples.



(b) Histogram of minimum values less than 10.

Figure 2.4: minimum value distributions.

A natural first step is to analyze the light curves for which the minimum χ^2 value is less than 10. In these cases, the shapelet is visually well-aligned when superposed on the light curve:

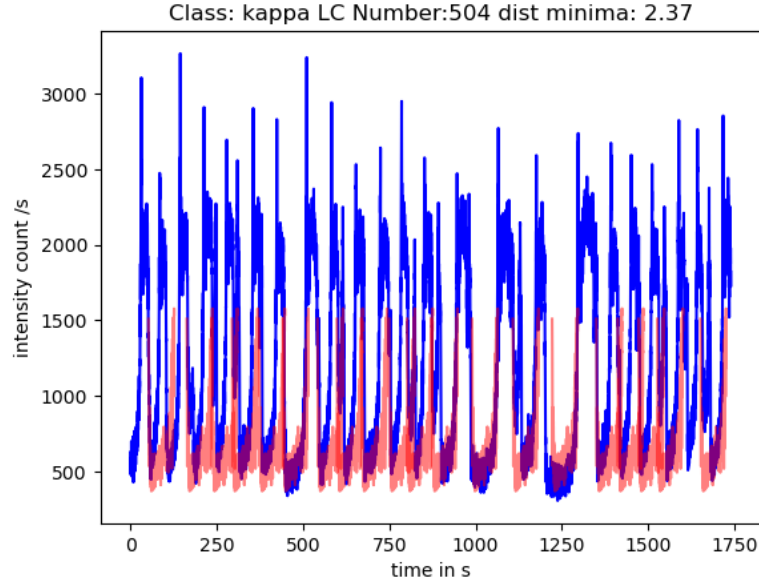


Figure 2.5: We see by superposing the shaplet on the lightcurve number 504 that the shaplet reproduces perfectly the lightcurve.

2.6 Finding a Cut and Efficient Cut Concept

The initial cut we considered for defining classes is in the range between 0 and 2.5, where the light curve and shapelet match almost perfectly, and beyond which noticeable differences begin to emerge.

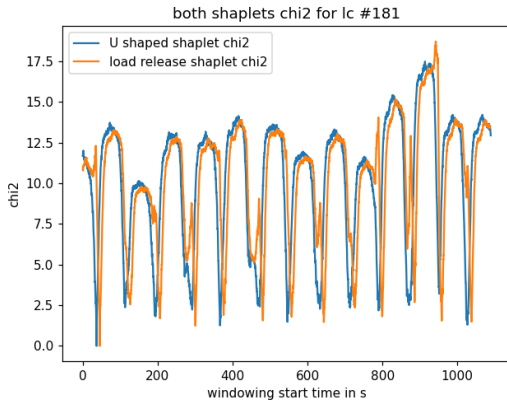
However, one consequence of applying the cut is that we cannot perfectly reconstruct the light curve, as there may be regions where the resemblance exceeds the 2.5 threshold. To address this, we introduce the concept of an *efficient cut*, defined as the minimal threshold at which a shapelet can continuously represent a light curve without any *free regions*—that is, contiguous gaps longer than twice the length of the shapelet.

Since our primary interest lies in the minimum points, we also examine the maximum of the minimum χ^2 values across the light curve. This leads us to the summary table, which includes the minimum χ^2 , the efficient χ^2 , and the maximum χ^2 for each curve.

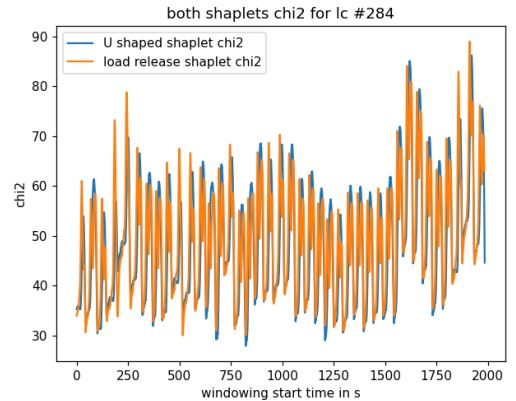
we see that for rho class, we can define hence two sub classes one from the lightcurve 181 to lightcurve 283 and another from 283 to the last lightcurve in rho class.

2.7 Robustness Testing

Once we validated the analysis using the U shapelet with lambda shapelet, we proceeded to the next step toward building a library of shapelets that are most representative of our light curves. The next shapelets we examined are the lambda shapelets. As shown by the selected summary statistics, the results are consistent with those obtained using the U shapelet. This consistency is also evident from the corresponding χ^2 distributions.



(a) Plot of the χ^2 values for both U shapelet and lambda shaplet applied to LC 181



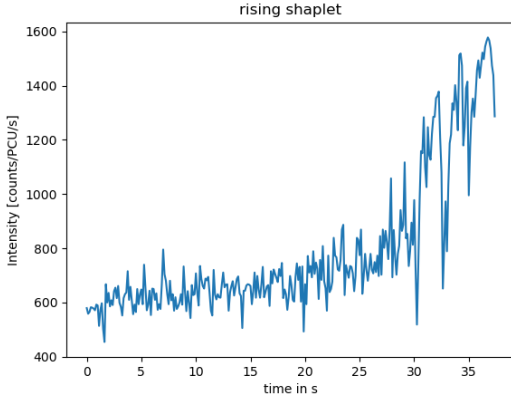
(b) Plot of the χ^2 values for both U shapelet and lambda shaplet applied to LC 284

Figure 2.6: The plots demonstrate that the results for U and lambda shapelets are consistent and coherent.

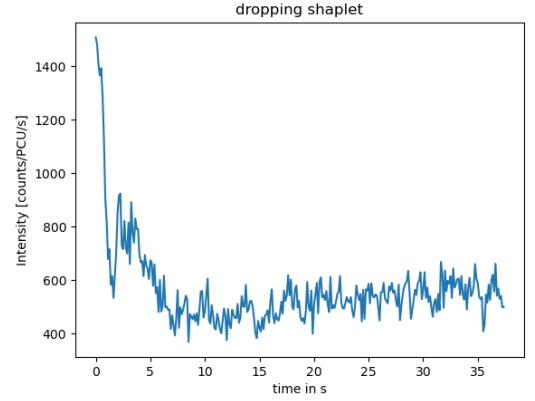
However, it is important to note that the 75-second shapelets often contain multiple modes. For instance, the lambda shapelet can be decomposed into a rising mode and a falling mode and that the. This observation allows us to define more targeted shapelets that isolate each mode individually. When comparing these refined shapelets to the light curves, we can extract segments corresponding specifically to rising and falling phases.

Additionally, we observed that partial match is considered a good match ('overall' the matching is close). Therefore, from this point on, we focus on shapelets of 32.5 seconds in length (300

points), as they provide more elementary and interpretable representations. As seen in the table below, these shorter shapelets also yield comparable accuracy.



(a) rising shaplet an example of subsequence of 300 points



(b) dropping shaplet an example of subsequence of 300 points

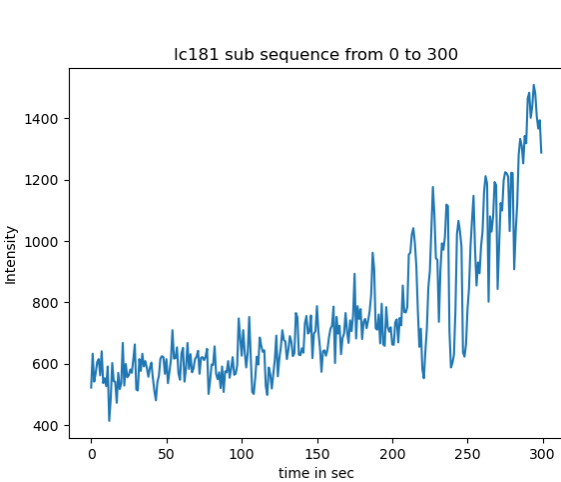
Figure 2.7: The plots depict 300-point shapelets that clearly reflect key physical processes in accretion: the mass buildup phase, its duration, and the sudden radiation outburst.

2.8 Averaging

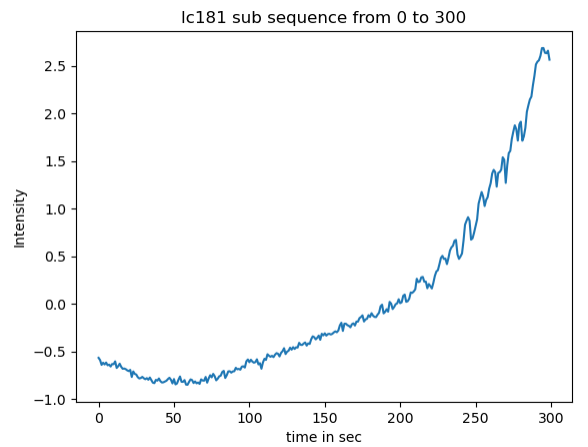
The idea behind averaging is to obtain a shapelet that is as representative as possible of the underlying pattern, while minimizing the influence of noise. To achieve this, we use the following formula:

$$\langle shaplet, \{S \subseteq L \mid \chi_{res}^2(S, shaplet) < 2\} \rangle$$

As a result, we obtain:



(a) Plot of a non averaged shaplet



(b) Plot of averaged shaplet with a z normalization

Figure 2.8: Plots showing the result of averaging

2.9 Scaling and normalisation

Another important observation is that some light curves appear to be perfectly periodic and nearly indistinguishable from the reference light curve, yet differ either by a vertical shift in intensity (i.e., a displacement along the y-axis), or by having a wider cycle (i.e., lower frequency).

We initially attempted to account for this by applying a simple scaling transformation; multiplying by a constant and adding an offset. However, this approach proved inadequate, particularly during the averaging process. When scaling a nearly flat, noise-dominated subsequence and a light curve independently, the standard deviation of the subsequence (being much lower) leads to mismatches in the comparison. As the additive offset increases, the shapelet diverges further from its intended form, ultimately resulting in a completely different pattern.

To overcome this issue, we adopted a different strategy: normalizing both the shapelets and the light curves before performing any comparisons. In this transformed space, the formula for the chi-squared metric becomes:

$$\begin{aligned}\alpha_{\text{shaplet}} &= \frac{1}{\sigma_{\text{shaplet}}} \\ \beta_{\text{shaplet}} &= -\frac{\langle \text{shaplet} \rangle}{\sigma_{\text{shaplet}}} \\ \alpha_{\text{lc}} &= \frac{1}{\sigma_{\text{lc}}} \\ \beta_{\text{lc}} &= -\frac{\langle \text{lc} \rangle}{\sigma_{\text{lc}}} \\ \chi^2 / \text{dof} &= \sum_{S \subseteq \text{lc}} \frac{(\alpha_{\text{lc}} \cdot S + \beta_{\text{lc}} - (\alpha_{\text{shaplet}} \cdot \text{shaplet} + \beta_{\text{lc}}))^2}{\alpha_{\text{lc}}^2 \cdot \langle S \rangle + \alpha_{\text{shaplet}}^2 \cdot \langle \text{shaplet} \rangle}\end{aligned}$$

Conclusion

Having confirmed the effectiveness of the shapelet approach for analyzing the lightcurves of GRS 1915+105, we are now well-positioned to construct a representative set of shapelets tailored to each variability class.

Chapter 3

Towards a Dictionary of Shaplets

3.1 Shaplet Selection and Recursive Sampling Strategy

While we have strong evidence supporting the effectiveness of our method, we have not yet resolved the issue of shaplet selection. All shaplets used thus far were chosen based on prior knowledge of the ρ class, and hence, are not agnostic to the dataset. From a preliminary investigation of the ρ class, we know that it exhibits perfect or near-perfect periodicity. Consequently, we selected patterns that recur in all light curves of the ρ class, modulo variations in frequency, intensity, and time scale; we also have not addressed the problem of anomaly detection.

To approach the shaplet selection problem, we propose the following idea: sample a random subsequence of 300 points. This random subsequence acts as a *seed* for a recursive algorithm designed to traverse the dataset intelligently and converge toward the optimal shaplets for classification. For now, we restrict the study to the ρ class. If the proof of concept proves successful, we plan to generalize this approach to other variability classes; a part of the work currently still in progress.

Algorithm Description:

1. Randomly choose an initial shaplet.
2. For every local minimum position in the lightcurve, take the interval between this minimum and the next one.
3. Check whether there is sufficient space for analysis: we require at least double the length of the shaplet (37.5×2 seconds), since the initial shaplet already occupies 37.5 seconds.
4. Within this interval, identify the position of the local maximum between $[\text{minpos} + 37.5, \text{nextpos}]$.
5. For each valid candidate:
 - Compute the χ^2 value of the alignment.

- Store this value in a sorted dynamic array with binary search-based insertion (using Python's `bisect` module).
 - Save the corresponding subsequence in a simple list of shaplets with index matching the value of its χ^2 in the sorted dynamic array.
6. After processing, extract the maximally different shaplet which is of index -1. Average it using the Averging process mentioned in the previous chapter.
 7. Repeat the process recursively. In the subsequent iterations, we construct a single curve by taking the pointwise minimum across all previously computed χ^2 curves (i.e., for each time index, take the minimum χ^2 value across all shaplets). The same algorithm is then applied again, but this time on the $\min(\chi^2)$ curve, treating it as the new chi2 to take from the candidate shaplets.
 8. Stop the recursion whenever there is no more free space.

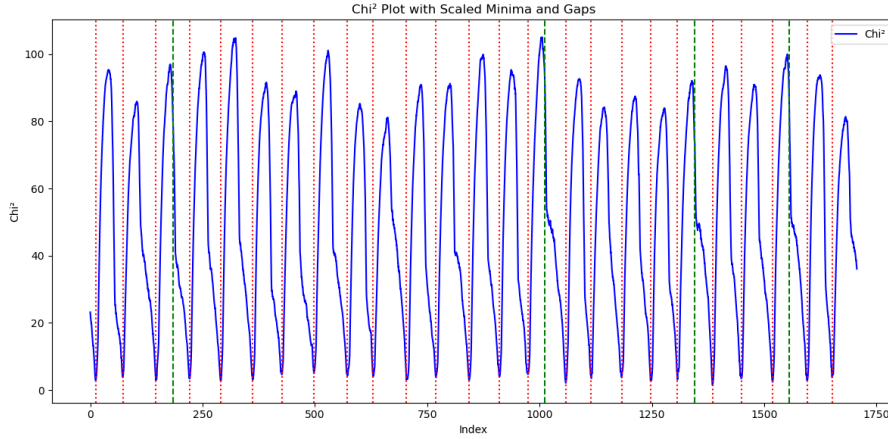


Figure 3.1: This image displays a χ^2 plot. The red lines mark the positions of each minimum. Gaps without green lines indicate short intervals, while the presence of a green line marks the start of a search window, extending up to the next red line.

Observations and Issues

Initial tests in the normalized space yield promising results. Across three runs, we observe the emergence of the desired structure: one rising component, one falling component, and one stable component. However, when the number of runs increases, we notice a troubling pattern. The χ^2 value of the newly generated shaplets start to oscillate rather than decrease monotonically. This suggests a lack of convergence and points to an area requiring further investigation. Ideally, we expect the maximum χ^2 value between all selected shaplets to decrease steadily with each iteration.

Conclusion and Future Perspectives

The long-term objective of this work is to develop a method for encoding lightcurves as sequences of discrete shapelets. For example, given a set of shapelets labeled a , b , and c , a lightcurve segment spanning five times the length of a shapelet can be symbolically represented as a sequence of five letters, each corresponding to the best-matching shapelet. This compact symbolic encoding paves the way for effective anomaly detection by analyzing patterns and deviations within these shapelet sequences.

Looking ahead, an exciting avenue is to use shapelets to characterize the timing features of the ρ class in a generalizable manner. In particular, studying the relationship between the timing and the height of peaks holds significant potential for advancing physical models of the ρ class, contributing deeper insights into the underlying astrophysical processes.

Bibliography

- [1] A. J. Castro-Tirado, S. Brandt, and N. Lund. “GRS 1915+105”. In: 5590 (Aug. 1992), p. 2.
- [2] R. P. Fender et al. “MERLIN observations of relativistic ejections from GRS 1915+105”. In: 304.4 (Apr. 1999), pp. 865–876. DOI: 10.1046/j.1365-8711.1999.02364.x. arXiv: astro-ph/9812150 [astro-ph].
- [3] Benjamin J. Ricketts et al. “Mapping the X-ray variability of GRS 1915 + 105 with machine learning”. In: 523.2 (Aug. 2023), pp. 1946–1966. DOI: 10.1093/mnras/stad1332. arXiv: 2301.10467 [astro-ph.HE].
- [4] T. M. Tauris and E. P. J. van den Heuvel. “Formation and evolution of compact stellar X-ray sources”. In: *Compact stellar X-ray sources*. Ed. by Walter H. G. Lewin and Michiel van der Klis. Vol. 39. 2006, pp. 623–665. DOI: 10.48550/arXiv.astro-ph/0303456.