



Universiteit  
Leiden  
The Netherlands

## Modelling variability power spectra of active galaxies from irregular time series

Lefkir, M.; Vaughan, S.; Huppenkothen, D.; Uttley, P.; Anilkumar, V.

### Citation

Lefkir, M., Vaughan, S., Huppenkothen, D., Uttley, P., & Anilkumar, V. (2025). Modelling variability power spectra of active galaxies from irregular time series. *Monthly Notices Of The Royal Astronomical Society*, 539(2), 1775-1795. doi:10.1093/mnras/staf532

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4292213>

**Note:** To cite this publication please use the final published version (if applicable).

# Modelling variability power spectra of active galaxies from irregular time series

Mehdy Lefkir<sup>1</sup>,<sup>\*</sup> Simon Vaughan<sup>1</sup>, Daniela Huppenkothen<sup>2,3</sup>, Phil Uttley<sup>3</sup> and Vysakh Anilkumar<sup>4</sup>

<sup>1</sup>*School of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK*

<sup>2</sup>*Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands*

<sup>3</sup>*SRON Netherlands Institute for Space Research, Niels Bohrweg 4, NL-2333 CA Leiden, the Netherlands*

<sup>4</sup>*Leiden Observatory, Leiden University, Postbus 9513, NL-2300 RA Leiden, the Netherlands*

Accepted 2025 March 28. Received 2025 March 24; in original form 2025 January 10

## ABSTRACT

A common feature of active galactic nuclei (AGNs) is their random variations in brightness across the whole emission spectrum, from radio to  $\gamma$ -rays. Studying the nature and origin of these fluctuations is critical to characterizing the underlying variability process of the accretion flow that powers AGNs. Random timing fluctuations are often studied with the power spectrum; this quantifies how the amplitude of variations is distributed over temporal frequencies. Red noise variability – when the power spectrum increases smoothly towards low frequencies – is ubiquitous in AGNs. The commonly used Fourier analysis methods, have significant challenges when applied to arbitrarily sampled light curves of red noise variability. Several time-domain methods exist to infer the power spectral shape in the case of irregular sampling but they suffer from biases which can be difficult to mitigate, or are computationally expensive. In this paper, we demonstrate a method infer the shape of broad-band power spectra for irregular time series, using a Gaussian process regression method scalable to large data sets. The power spectrum is modelled as a power-law model with one or two bends with flexible slopes. The method is fully Bayesian and we demonstrate its utility using simulated light curves. Finally, Ark 564, a well-known variable Seyfert 1 galaxy, is used as a test case and we find consistent results with the literature using independent X-ray data from *XMM-Newton* and *Swift*. We provide publicly available, documented, and tested implementations in PYTHON and JULIA.

**Key words:** methods: numerical – methods: statistical – galaxies: active – galaxies: individual: Ark 564 – X-rays: galaxies.

## 1 INTRODUCTION

An active galactic nucleus (AGN) is a bright region located at the centre of many galaxies, powered by accretion of gas on to a supermassive black hole (Padovani et al. 2017). Observations of the central engine of AGN will contain information about the physics of mass accretion around a supermassive black hole, but understanding how to interpret the observed patterns and use them to study the accretion flow and black hole remains an open challenge. The black hole mass is typically in the range  $10^6 - 10^9 M_\odot$ , and the accretion process can produce radiation from radio to X-rays as well as long-lasting relativistic jets (Rees 1984). In some cases, it can even generate powerful winds that affect the host galaxy evolution (Fabian 2012). The central engine is usually too small to be resolved, the exceptions are the recent radio observations with the Event Horizon Telescope that resolve the vicinity around the black hole in M87 and in the galactic centre (Event Horizon Telescope Collaboration 2019, 2022). In all other cases, information about the physics and geometry of the central engine must be extracted from the light

received: its spectrum, variability through time, and its polarization (or a combination of these).

A feature of many AGNs is their variability in brightness. In almost all cases the variations appear random, without predictable trends or cycles, with larger amplitude variations on longer time-scales. Although the variations are random, we typically assume that they are drawn from a process which is *weakly stationary*. By *stationary* we mean the statistical properties of the process remain constant over time. By *weakly stationary* we mean that the first two *moments* of the random process do not change with time: the mean and the autocovariance. The Fourier transform of the autocovariance of a random process is known as the power spectrum, it contains the same information but for many problems it is simpler to work with than the autocovariance. Weak stationary implies a time-constant power spectrum, where the power spectrum determines the contribution to the total variance of the process from random fluctuations on different time-scales (inverse temporal frequencies).

AGN variability is often described as ‘red noise’ (Marshall, Warwick & Pounds 1981; Papadakis & Lawrence 1993), meaning that the power spectrum of the variations rises towards low temporal frequencies, approximately following a power-law form  $f^{-\alpha}$ , with power-law slope  $\alpha > 1$  (and typically  $\alpha > 2$ ) at high frequencies. In order for the total variance of the process to remain finite, the integral

\* E-mail: [ml556@leicester.ac.uk](mailto:ml556@leicester.ac.uk)

of the power spectrum over all frequencies must converge; this requires that the power spectrum at low frequencies has a slope  $\alpha < 1$  (Press 1978). The frequency of this flattening defines a characteristic time-scale, presumably set by the size and other properties of the central engine, and can be used to compare different AGNs. In most cases, measuring this time-scale requires long, months–years duration monitoring campaigns consisting of light curves, which are sparsely and often irregularly sampled. Accurate power spectral measurement then requires the use of methods that can account for the effects of this sampling.

In X-ray astronomy, the standard approach is to work in the Fourier domain by estimating the power spectrum using the periodogram (van der Klis 1989; Bachetti & Huppenkothen 2022). This is most naturally applied to regularly sample time series data; one way to handle the sparse and irregular sampling of real data is to use forward-modelling. Here, simulations of light curves drawn from a process with a given power spectral shape are resampled to match the observed sampling pattern (e.g. Uttley, McHardy & Papadakis 2002), and thereby include in the modelling process any distortions arising from the time sampling. Using these methods, X-ray power spectral bends or breaks have been reported for many AGNs (Edelson & Nandra 1999; Uttley et al. 2002; Markowitz et al. 2003a), showing that the bend time-scale linearly scales with black hole mass. This  $t_{\text{bend}} \propto M_{\text{BH}}$  relation can be extrapolated to match similar power spectral bends in stellar mass black hole X-ray binaries (McHardy et al. 2006). The comparison with black hole X-ray binaries (BHXRBs) also makes the low-frequency power spectral shapes of interest. For example most AGNs show single power spectral bends from high-frequency slopes  $\alpha > 2$  to low-frequency slopes  $\alpha \simeq 1$ , similar to those seen in BHXRB soft states (McHardy et al. 2004; Uttley & McHardy 2005). A few AGNs show evidence for a second, lower frequency bend, marking a similarity with BHXRB hard and hard-intermediate states (McHardy et al. 2007; Summons et al. 2007). More generally, the broad-band power spectral shapes of accreting black holes can be used to constrain the physical origin of the variability process, for example in terms of accretion rate variations produced at different radii in the accretion disc (Arévalo & Uttley 2006; Uttley & Malzac 2024).

The requirement of regular time sampling for the conventional periodogram means its application to AGN data is limited, since such data are usually irregularly sampled. Furthermore, the periodogram is subject to Fourier leakage effects – undesirable distortions and correlated errors between sampled frequencies. For example steep power spectral slopes at low frequencies lead to so-called ‘red noise leak’ which contributes a bias with slope  $\alpha = 2$  (Uttley et al. 2002) and makes it difficult to accurately constrain steeper power spectral slopes.

Fourier methods using forward-modelling approaches have been proposed to overcome these problems (Done et al. 1992; Uttley et al. 2002). In these approaches, the data are interpolated on to an even grid and the periodogram is computed. This is then compared to the results of simulations based on a model power spectrum: long simulations of time series are generated, resampled to match the observed data, and the interpolation and periodogram computations applied to the simulation. This process is repeated over many simulations to produce an average periodogram to be compared with the data. These methods allow for analysis of irregularly sampled data but, being simulation-based, the estimation of the best-fitting power spectral shape is slow and limited to only a few free parameters. The likelihood function for this is not known, and so a fit statistic is used for its similarity to standard chi-square fitting. The distribution of the fit statistic is not known; quantifying the goodness of fit and

parameter uncertainties requires a large number of simulations, and depends on the choice of fit statistic (see Summons et al. 2007; Mueller & Madejski 2009; Markowitz 2010, for a discussion of some issues).

In optical AGN variability studies, sampling limitations can be even more significant, for example due to weather, telescope scheduling, and seasonal visibility of targets from ground-based observatories. However, unlike X-ray astronomy, optical studies of AGN variability often work in the time domain (e.g. Zu et al. 2013). Most time domain methods have the advantage of being free of the biases of Fourier forward-modelling of the power spectrum, but at the expense of assuming a statistical distribution (and hence power spectral form) for the process (Scargle 1981), which may make the method unsuitable for modelling observed AGN light curves (e.g. see Kozłowski 2016). However, these methods are not biased by irregular sampling and can account for heteroscedasticity when the error bars are different between flux measurements. Moreover, if the likelihood can be calculated in terms of the fitted model parameters, rapid gradient-descent-based fitting approaches are possible which enable rapid modelling of large samples of light curves, for example from massive time-domain surveys.

Over the years, Gaussian processes (GPs) have been widely adopted in time domain astronomy (Aigrain & Foreman-Mackey 2023) and applied to study time series. Using standard GP regression, Miller et al. (2010) and Zoghbi, Reynolds & Cackett (2013) provided first attempts at modelling flexible power spectra but were computationally expensive. The damped random walk (DRW), a particularly simple GP model, has recently become popular as a model for quasar variability (see e.g. Kelly, Bechtold & Siemiginowska 2009). The power spectrum of a DRW process is a Lorentzian centred at zero frequency, with a high-frequency slope  $\alpha = 2$ , inconsistent with many studies of AGN power spectra in both X-ray (González-Martín & Vaughan 2012) and optical bands (Edelson et al. 2014).

Kelly et al. (2014) introduced continuous autoregressive moving average (CARMA) processes to astronomy. These enable a fast and flexible way of modelling stochastic time series. This power spectrum model can be seen as a weighted sum of modified Lorentzians (Anilkumar 2024). Although this is more flexible than a single Lorentzian, the power spectrum decays with a fixed integer slope depending on the order of the process and flattens at low frequencies. It can also model narrow features such as quasi-periodic oscillations (QPOs) on top of the broad-band power spectrum. The parametrization of this model makes it difficult to properly define the relatively featureless power spectra expected in most models of AGN variability. Proper characterization of the continuum noise is crucial for defining a null hypothesis when quantifying the significance of periodicities or QPOs in red noise (see Vaughan 2010; Anilkumar 2024). A limitation of CARMA models for this purpose is that they can implicitly include QPO-like signals in the underlying power spectrum.

A sophisticated GP time domain model named *celerite* was presented in Foreman-Mackey et al. (2017). This model allows for steep power spectra in the form of  $f^{-4}$  with a fixed integer slope at high frequencies and fast inference. This later tool is widely used for exoplanet science (e.g. Espinoza, Kossakowski & Brahm 2019), and has been used occasionally for variability studies of active galaxies (e.g. Zhang et al. 2021; Zhang, Yang & Dai 2023). For all of these models, bend frequencies may be flexible but the high-frequency slope is always a fixed integer – either  $-2$  or  $-4$  and the low-frequency power spectrum is constrained to be flat.

In this paper, we present a new GP model which aims to avoid all these limitations – a method that will work on data with arbitrary

time sampling, made up of data points which may have very different measurement uncertainties and perhaps even come from different instruments, but allows the use of models for the power spectrum shape that are consistent with our best current understanding of AGN variability. This uses GP models for their flexibility and ability to model irregular data, providing a well-defined likelihood function that can be used in Bayesian modelling. Our innovation here is to combine the GP framework with a way to approximate the desired power spectral shapes in a way that allows for fast computation. Although developed specifically to aid the study of AGN variability, it can be applied to time series data in general. We make use of flexible bending power-law models for the power spectrum by approximating the model with simple basis functions chosen for their amenability to fast computational methods. This allows proper estimation of the bend frequencies and slopes of the power spectrum with a simpler parametrization compared to current methods. Building the method with the basis functions of Foreman-Mackey et al. (2017) allows fast likelihood computation on large data sets.

In Section 2, we define formally the power spectrum and the statistical assumptions of this work in the context of astronomical time series and present GP regression in the general context and current models used in astronomy. Section 3 presents our approach to the power spectrum approximation and how we implement an efficient likelihood evaluation. In Section 4, we present the Bayesian workflow adopted in this work, from the modelling to diagnostics post-inference. We test our method with simulation-based calibration in Section 5. We apply our method to the long-term light curve of the Seyfert 1 galaxy Ark 564 in Section 6 to estimate its power spectrum. We discuss limitations and possible improvements of this method in Section 7. Section 8 concludes on the method and results presented in this work.

In order to describe the data, models, associated statistics and their properties, we require a system of notation. Here, we adopt notation for the GP theory drawn from Rasmussen & Williams (2006). Specifically, a subscript asterisk is used to denote a test set, for example  $\mathbf{t}$  might be a vector of times where a process has been observed, and  $\mathbf{t}_*$  is a set of times at which we wish to predict the value of the process. Vector quantities are in bold (e.g.  $\mathbf{x}$ ; a column vector), matrices are uppercase Roman or Greek letters (e.g.  $K$ ,  $\Sigma$ ).  $K^T$  and  $K^{-1}$  are the transpose and inverse of the matrix  $K$ , respectively, and  $|K|$  is its determinant. We use  $\sim$  to mean ‘distributed according to’ to indicate the probability distribution of a variable or process. Table 1 summarizes the notation and symbols used in this paper.

## 2 RANDOM PROCESSES

If a random process  $\{X(t)\}$  is weakly stationary, also known as *wide sense stationary* (WSS), it can be fully described by its mean  $\mu_X$  and autocovariance function  $\mathcal{R}(\tau)$  given by equation (1). The expectation operator  $\mathbb{E}$ , represents the ensemble mean, the average over all possible realizations of the process  $\{X(t)\}$ , this is not the time average. From now on, we will assume to be working with random processes that are stationary in the wide sense.

$$\begin{aligned}\mathcal{R}(\tau) &= \text{Cov}(X(t), X(t + \tau)) \\ &= \mathbb{E}[(X(t) - \mu_X)(X(t + \tau) - \mu_X)].\end{aligned}\quad (1)$$

The autocovariance function of the process quantifies how the values of the time series are correlated with each other when separated in time by  $\tau$ . It is worth noting that the variance of the process is given by  $\mathcal{R}(0) = \text{Var}(X)$  and should always be positive. The Wiener–Khinchin theorem states that the autocovariance function and the

**Table 1.** Symbols used throughout this work.

|                                              | Observation                                        |
|----------------------------------------------|----------------------------------------------------|
| $\mathbf{x}$                                 | Observed time series                               |
| $\mathbf{t}$ or $t$                          | Time                                               |
| $\sigma^2$                                   | Variance of the measurement process                |
| $\Delta t$                                   | Sampling period or time spacing in the time series |
| $T$                                          | Duration of the time series                        |
| $f_{\min} = 1/T$                             | Minimum frequency in the time series               |
| $f_{\max} = 1/2\min(\Delta t)$               | Maximum frequency in the time series               |
|                                              | Gaussian process                                   |
| $f$                                          | Gaussian process                                   |
| $f_* \mathbf{t}, \mathbf{x}, \mathbf{t}_*$   | Conditioned Gaussian process                       |
| $\mathbf{t}_*$                               | Times for the prediction                           |
| $\mathbf{x}_*(\mathbf{t}_*)$                 | Predicted time series                              |
|                                              | Modelling                                          |
| $\mathcal{R}(\tau)$                          | Autocovariance function                            |
| $\mathcal{P}(f)$                             | Power spectral density                             |
| $f$                                          | Frequency                                          |
| $\tau$                                       | Time delay                                         |
| $S_{\text{low}}$                             | Scale factor to extend the low-frequencies         |
| $S_{\text{high}}$                            | Scale factor to extend the high-frequencies        |
| $f_{\text{start}} = f_{\min}/S_{\text{low}}$ | Minimum frequency                                  |
| $f_{\text{stop}} = f_{\max}S_{\text{high}}$  | Maximum frequency                                  |
| $J$                                          | Number of basis function for the approximation     |
| $\psi(f)$                                    | Basis function in the Fourier domain               |
| $\phi(\tau)$                                 | Basis function in the time domain                  |
|                                              | Parameters                                         |
| $\theta$                                     | Parameters                                         |
| $\alpha_i$                                   | Index/slope of the power law                       |
| $f_{b,i}$                                    | Bend frequency                                     |
| variance                                     | $\mathcal{R}(0)$ or integral of $\mathcal{P}(f)$   |
| $\nu$                                        | Scale factor on the measurement uncertainties      |
| $\mu$                                        | Mean of the Gaussian time series                   |
| $c$                                          | Constant to shift a log-normal time series         |
| $\gamma$                                     | Inter-calibration factor between two time series   |
|                                              | Inference                                          |
| $\mathbf{d}$                                 | Data                                               |
| $p(\mathbf{d} \theta) = \mathcal{L}(\theta)$ | Likelihood function                                |
| $p(\theta)$                                  | Prior probability density                          |
| $Z$                                          | Bayesian evidence                                  |
| $BF$                                         | Bayes factor                                       |

power spectrum for frequency  $f$ ,  $\mathcal{P}(f)$ , are Fourier pairs (Priestley 1981; Chatfield 2004):

$$\mathcal{P}(f) = \int_{-\infty}^{+\infty} \mathcal{R}(\tau) e^{-2i\pi f\tau} d\tau, \quad \mathcal{R}(\tau) = \int_{-\infty}^{+\infty} \mathcal{P}(f) e^{2i\pi f\tau} df. \quad (2)$$

For this work, we will consider power spectral densities which yield finite variance and thus must be absolutely integrable. By definition, if the power spectrum is well-behaved, the autocovariance function of the process will be positive semidefinite.

### 2.1 Gaussian processes

Here, we introduce GPs which enable the modelling of the variability of Gaussian time series with arbitrary sampling using autocovariance function. Over the last decade, GPs have been widely used in time domain astronomy with many robust and well-tested codes – see the review by Aigrain & Foreman-Mackey (2023).

GPs are a class of random processes for which the joint probability distribution of any finite set of random variables is a Gaussian distri-



bution (Rasmussen & Williams 2006). The probability density of a multivariate ( $D$ -dimensional) Gaussian with mean  $\mu$  and covariance matrix  $\Sigma$  is given by equation (3).

$$p(\mathbf{x}|\mu, \Sigma) = (2\pi)^{-D/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right). \quad (3)$$

A GP  $f$ , with one-dimensional output, is described by a mean function  $\mu(t)$  and a covariance function  $k(t, s)$ . We will assume the mean function to be constant  $\mu(t) = \mu$ . As we intend to infer the statistical properties of a time series modelled with a time-symmetric stationary process, the covariance function will depend only on the time separation  $\tau = |t - s|$ . We write equivalently  $\mathcal{R}(\tau) = k(t, t + \tau)$ .

### 2.1.1 Inference with Gaussian processes

When the covariance function is a function of a vector of parameters  $\theta$ , then we can find the set of ‘best-fitting’ parameters to perform regression. This can be done by either maximizing the log-likelihood or in a Bayesian framework sampling a posterior distribution. The log-likelihood function associated with the GP is given in equation (4).

$$\ln \mathcal{L}(\theta, \nu, \mu) = -\frac{1}{2}(\mathbf{x} - \mu)^T (K + \nu \sigma^2 I)^{-1}(\mathbf{x} - \mu) - \frac{1}{2} \ln |K + \nu \sigma^2 I| - \frac{N}{2} \ln(2\pi). \quad (4)$$

where  $\mathbf{x}$  is the vector containing the noisy observed time series with measurement variance  $\sigma^2$  at times  $\mathbf{t}$  and  $I$  is the identity matrix.  $K = k(\mathbf{t}, \mathbf{t})$  is the covariance matrix obtained by evaluating the covariance function at the observed times.  $\mu$  is the mean of the time series,  $\nu$  is a scaling parameter on the measurement variance  $\sigma^2$ . If  $\sigma^2$  is a good approximation to the variance of the measurement process  $\nu$  should be around one, i.e. the error bars are reliable. If  $\nu > 1$  then the uncertainties are underestimated, alternatively if  $\nu < 1$  the uncertainties are overestimated (see e.g. Vaughan et al. 2016). The first term in the log-likelihood can be seen to improve the quality of the fit, the second term penalizes complex models and the last term is a normalizing constant.

### 2.1.2 Prediction

Given a set of parameters  $\theta$ , we can also predict the time series at observed and arbitrary times. A GP can be used to perform regression on a time series  $\mathbf{x}$ , regardless of the sampling, to proceed, one must first choose the covariance function. Gaussian distributions have several convenient properties, the marginalized and conditioned distributions of Gaussian random variables are also Gaussian. Therefore, GP regression can be seen from a conditional probability point of view as the posterior distribution obtained after conditioning a prior distribution with the data. The prior GP is only defined by the choice of mean  $\mu(t)$  and covariance function. The posterior or conditioned process  $f_*$  is Gaussian with mean and covariance matrix given by equations (5) and (6).

$$\mathbb{E}[f_*|\mathbf{t}, \mathbf{x}, \mathbf{t}_*] = K_* [K + \sigma^2 I]^{-1} \mathbf{x} \quad (5)$$

$$\text{Cov}[f_*|\mathbf{t}, \mathbf{x}, \mathbf{t}_*] = K_{**} - K_* [K + \sigma^2 I]^{-1} K_*^T. \quad (6)$$

Where  $\mathbf{x}$  is the vector containing the noisy observed time series with measurement variance  $\sigma^2$  at times  $\mathbf{t}$  and  $I$  is the identity

matrix.  $K = k(\mathbf{t}, \mathbf{t})$  is the covariance matrix obtained by evaluating the covariance function at the observed times,  $K_{**} = k(\mathbf{t}_*, \mathbf{t}_*)$  is the equivalent for the prediction times  $\mathbf{t}_*$ .  $K_* = k(\mathbf{t}_*, \mathbf{t})$  is the covariance matrix between observed and prediction times. Adding the term  $\sigma^2$  to the diagonal of the covariance matrix assumes that all values of  $\sigma^2$  are independent and Gaussian-distributed. The vector of measurement errors  $\sigma^2$  can account for the individual measurements having different sizes of error bar (i.e. heteroscedastic data).

It is worth noting that the computation of the posterior distribution in equations (5) and (6) and likelihood in equation (4) contains a matrix inversion. In practice, a Cholesky decomposition is used to compute this term but the computational cost of this operation scales as  $\mathcal{O}(N^3)$  where  $N$  is the number of points in the time series. Additionally, the cost for storing the covariance matrix is  $N \times N$  which makes regression difficult on large data sets – when  $N > 1000$ .

## 2.2 Covariance functions

The choice of covariance function (also sometimes called ‘kernel’ in the literature) will directly impact the shape and smoothness of the realizations of the GP (Rasmussen & Williams 2006). To be properly defined, the covariance function must be positive definite, which means it yields a positive definite covariance matrix. By definition, if the power spectral density is positive-valued, the covariance function is positive definite. As stated earlier, GP regression is generally limited to a small number of points but several covariance functions are structured in a way that makes the computations tractable on large data sets. Here, we present the main covariance functions used for AGN time-series modelling and their associated power spectral forms.

### 2.2.1 Exponential

The exponential covariance function is associated with the Lorentzian power spectrum ( $\psi_2$  in Table 2). A GP with this covariance is often referred to as a DRW in astronomy or an Ornstein–Uhlenbeck process. The high-frequency power spectrum decays as  $1/f^2$  while the low-frequency power spectrum is flat.

As said earlier, GP regression is generally limited to a small number of points but several covariance functions are structured in a way that makes the computations tractable on large data sets.

### 2.2.2 CARMA

Based on the works of Jones (1981), Jones & Ackerson (1990) and Belcher, Hampton & Wilson (1994), Kelly et al. (2014) introduced CARMA processes to astronomy for inference of the power spectrum. A CARMA( $p, q$ ) process is composed of an autoregressive process of order  $p$  and a moving average process of order  $q$ , where  $0 \leq q \leq p - 1$ . This process is defined according to a stochastic differential equation (SDE) and in the assumption of Gaussian noise, it is a GP with analytical covariance function and power spectrum. The order ( $p, q$ ) of the SDE relates to the number of parameters and can be arbitrarily large, providing a very flexible modelling. The covariance function is a sum of complex exponentials and the power spectrum can be expressed as a weighted sum of  $p$  modified Lorentzians (Anilkumar 2024).

The cost of this method scales linearly with the number of data points, this comes from the state-space representation of the process which enables the use of Kalman recursions (Brockwell & Davis 2016). By choosing the order ( $p, q$ ) it is possible to infer the shape

**Table 2.** Basis functions and their autocovariance function for Gaussian process regression. **References:** (1) Rasmussen & Williams (2006), (2) Foreman-Mackey et al. (2017), and (3) this work.

| Name         | Basis function $\psi(f)$                      | Associated autocovariance function $\phi(\tau)$                                                                                                                                              | Ref. |
|--------------|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| DRW          | $\psi_2(f) = \frac{A}{\alpha^2 + 4\pi^2 f^2}$ | $\phi_2(\tau) = \frac{A}{2\alpha} \exp(-\alpha \tau )$                                                                                                                                       | (1)  |
| SHO          | $\psi_4(f) = \frac{1}{1 + f^4}$               | $\phi_4(\tau) = \frac{\pi}{\sqrt{2}} \exp(-\pi\sqrt{2} \tau ) \left( \cos(\pi\sqrt{2} \tau ) + \sin(\pi\sqrt{2} \tau ) \right)$                                                              | (2)  |
| DRW+Celerite | $\psi_6(f) = \frac{1}{1 + f^6}$               | $\phi_6(\tau) = \frac{\pi}{\sqrt{3}} \left( \frac{\exp(-2\pi \tau )}{\sqrt{3}} + \exp(-\pi \tau ) \left( \frac{\cos(\pi\sqrt{3} \tau )}{\sqrt{3}} + \sin(\pi\sqrt{3} \tau ) \right) \right)$ | (3)  |

of the power spectrum with a flexible sum of Lorentzians. One of the caveats of this method for estimating power spectra is that the number of parameters increases with the order so more parameters will need to be constrained which would make the inference slower.

### 2.2.3 Celerite

Foreman-Mackey et al. (2017) generalized the CARMA model to a broader class of covariance functions called *celerite*. This covariance function is a mixture of  $J$  exponentials and trigonometric functions with  $4J$  coefficients. With specific constraints on the coefficients, the power spectrum decreases as  $\propto f^{-4}$  at high frequencies and is flat at low frequencies. The *celerite* covariance matrix has a semiseparable structure (Ambikasaran 2015) which is exploited in a fast and stable algorithm presented in Foreman-Mackey et al. (2017). The GP log-likelihood and the posterior mean of the GP can be computed in a  $\mathcal{O}(NJ^2)$  time. Similarly, the memory cost of the operations is drastically reduced as the full covariance matrix is not saved in the memory. Finally, thanks to the convenient properties of semiseparable matrices, the product and the sum of *celerite* covariance functions are also semiseparable. A specific covariance function of interest for this work is the simple harmonic oscillator (SHO) covariance function associated with a stochastically driven damped simple harmonic oscillator. Its application is presented in Section 3.2.

## 3 PIORAN: A POWER SPECTRUM INFERENCE METHOD

We are interested in inferring the statistical properties of the underlying process generating the variability observed in AGN light curves. To do so, we want to estimate the power spectrum of AGNs. As Fourier methods can be biased and limited to regular data we will work in the time domain using GPs. We will make use of the covariance functions presented in the previous section to build a general and scalable method to infer the shape of the power spectrum of AGN light curves.

We present *pioran*,<sup>1</sup> a method to infer the broad-band power spectrum of random time series using GPs.

### 3.1 Model for the continuum variability

The continuum power spectrum of accreting black holes is well modelled with power laws and Lorentzians (e.g. McHardy et al. 2007). These phenomenological models are based on the examination of many periodograms of AGN and X-ray binary time series from previous work. In this work, we model the power spectrum with a

bending power-law model (McHardy et al. 2004; Summons et al. 2007), with  $n$  bends located at frequencies  $f_i$  and  $n + 1$  slopes  $\alpha_i$ .

$$\mathcal{P}(f) = A \left( \frac{f}{f_{b,1}} \right)^{-\alpha_1} \prod_{i=1}^n [1 + (f/f_{b,i})^{(\alpha_{i+1}-\alpha_i)}]^{-1}. \quad (7)$$

Here, we only use up to  $n = 1 - 2$  bends. When  $\alpha_1 \simeq 0$  and  $\alpha_2 \simeq 2$ , this model is similar to a Lorentzian centred at zero and the process reduces to a DRW. To have a finite integrated power, the low-frequency slope must be less than one, i.e.  $\alpha_1 < 1$  and the high-frequency slope  $\alpha_{n+1}$  must be steeper than one, i.e.  $\alpha_{n+1} > 1$ . Another way to ensure finite power is to use cut-offs to limit power at low and high frequencies. To use this flexible model in a GP framework, one has to compute the associated autocovariance function with the inverse Fourier transform. Unfortunately, there is no known analytical Fourier transform for such a model. We therefore cannot write a close form expression for the autocovariance function needed for equation (4). We can only rely on approximate methods. Here, we present a method for approximating the autocovariance function and using it in the GP regression.

### 3.2 The approximation

We approximate our power spectrum model,  $\mathcal{P}(f)$  in equation (7), using a finite set of simple functions as shown in equation (8).

$$\tilde{\mathcal{P}}(f) = \sum_{j=0}^{J-1} a_j \psi(f/f_j) \quad (8)$$

We call the functions  $\psi(f)$  ‘basis functions’, although these do not form a proper basis. They can approximate a certain range of power spectral shapes, but not all valid power spectra can be expressed as a sum of these functions. The inverse Fourier transform of these basis functions forms an approximation to the autocovariance model  $\mathcal{R}(\tau)$ .

$$\tilde{\mathcal{R}}(\tau) = \sum_{j=0}^{J-1} a_j f_j \phi(\tau f_j) \quad (9)$$

We use four requirements to inform the choice of basis functions,  $\psi(f)$ , they should:

- (i) be smooth in the frequency domain;
- (ii) have a power-law shape at lowest and highest frequencies;
- (iii) have a known autocovariance function; and
- (iv) have an autocovariance function amenable to fast computation methods.

Requirements (i) and (ii) ensure that the sum of a relatively lower number of basis functions can approximate well a smooth, broad-band spectrum with power law form at the highest and lowest frequencies. Requirement (iii) enables us to apply GP regression (equation 4), and requirement (iv) allows us to apply these methods

<sup>1</sup>Power spectrum Inference Of RANDOM time series.

efficiently. The *celerite* (Foreman-Mackey et al. 2017) covariance functions will satisfy these requirements. In Table 2, we present the basis functions  $\psi$  used for the approximation and their associated inverse Fourier transform  $\phi(\tau)$ .  $\psi_4$  is the power spectrum associated with a stochastically driven damped SHO when the quality factor equals  $Q = 1/\sqrt{2}$  (Foreman-Mackey et al. 2017).  $\psi_6$  can be obtained using partial fraction decomposition and recognizing it to be the sum of a Lorentzian and a general *celerite* power spectrum, this is detailed in Appendix A.

Assuming we can approximate our power spectrum model as in equation (8), the approximated autocovariance function is given by equation (9) where  $a_j$  are real coefficients and characteristic frequencies  $f_j$  of the basis functions.  $J$  is the number of basis functions. It should be stressed that due to the shape of the basis functions, the approximated model  $\tilde{\mathcal{P}}(f)$  will be flat at low frequencies – slope of 0 – and steep at high frequencies – either a slope of  $-4$  or  $-6$  depending on whether  $\psi_4$  or  $\psi_6$  was used. This implies that the integrated power (variance) will always be finite, without the need to introduce additional cut-offs.

The frequencies  $f_j$  are geometrically spaced and thus given by:  $f_j = f_{\text{start}} (f_{\text{stop}}/f_{\text{start}})^{j/(J-1)}$  for  $j = 0, 1, \dots, J-1$ . We introduce two scale factors  $S_{\text{low}}$  and  $S_{\text{high}}$  to extend the low and high frequencies from  $f_{\text{start}} = f_{\text{min}}/S_{\text{low}}$  to  $f_{\text{stop}} = f_{\text{max}} \cdot S_{\text{high}}$ . In practice, we choose  $S_{\text{low}} = 100$  and  $S_{\text{high}} = 20$ . The constraint  $\tilde{\mathcal{P}}(f_j) = \mathcal{P}(f_j)$ , forms a system of  $J$  linear equations where the unknowns are the  $a_j$ . The matrix  $B$  of dimensions  $(J, J)$  representing this system is presented in equation (10) and one can notice it has a Toeplitz form.

$$\mathbf{p} = \mathbf{aB} \quad \text{where } B_{ij} = \psi(f_i/f_j) \text{ and } p_j = \mathcal{P}(f_j) \quad (10)$$

Due to the small size of the  $B$  matrix  $J \leq 50$ , standard methods can be used to solve the system. In Appendix B1, we find that the matrix can be well-conditioned when the ratio of minimal and maximal frequencies spans several orders of magnitudes. Otherwise, one might need to use fewer basis functions. The approximation in the Fourier domain can be visualized in Fig. 1.

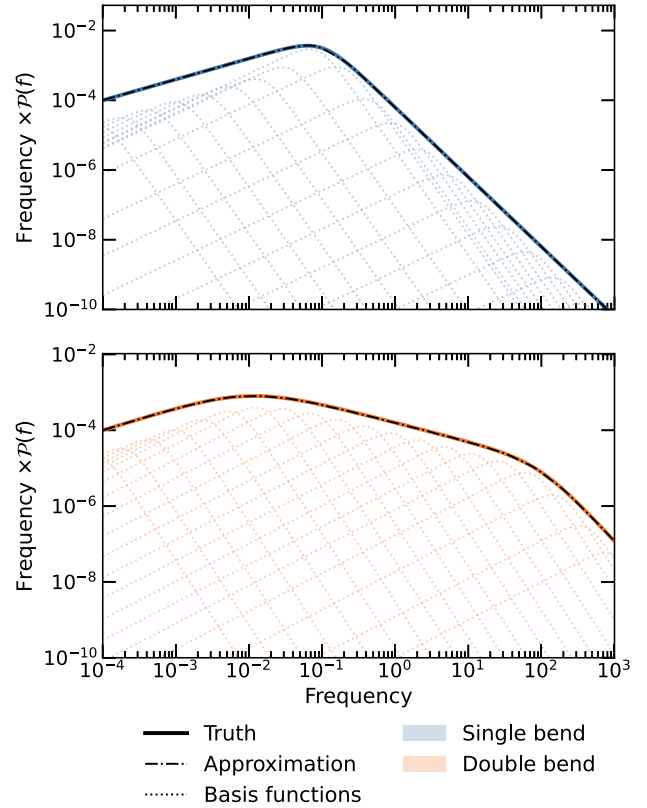
Once the  $a_j$  and  $f_j$  are obtained, they can be used in the *celerite* algorithm (Foreman-Mackey et al. 2017) to describe the covariance functions and perform regression with a computational time scaling linearly with the number of data points.

### 3.2.1 Checking the approximation

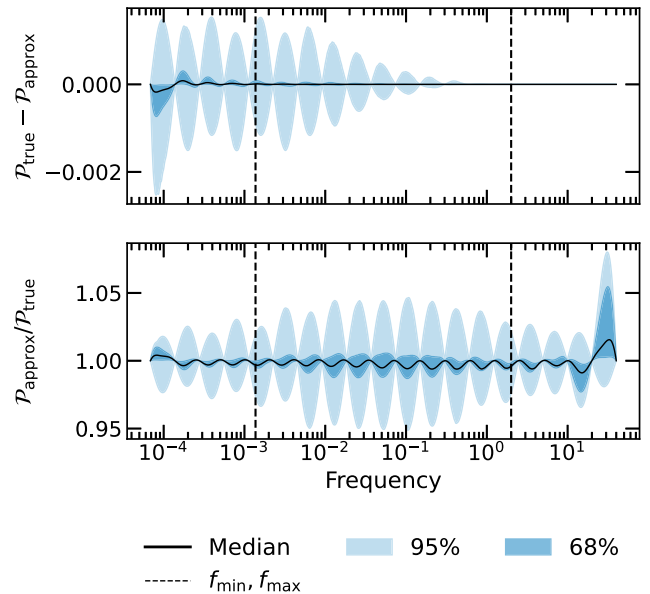
Before running any inference with this method, one should check that the approximation is accurately describing the model. To check this, we compute residuals and ratios between the true model  $\mathcal{P}$  and the approximated model  $\tilde{\mathcal{P}}$ .

First, we draw realistic values for the parameters of the intended model (bending power law), this can be done by sampling from a prior distribution (see Section 4.2.1 for more details about the priors). Then, for each set of parameters, we compute the power spectrum model and its approximation over a grid of frequencies to obtain a distribution of the intended and the approximated power spectral shapes. The distribution of the residuals and ratios are respectively given by the difference or ratio between  $\mathcal{P}$  and  $\tilde{\mathcal{P}}$ . We assess the quality of the approximation by computing quantiles on the distribution of the residuals and ratios; an example is shown in Fig. 2 where we approximate the model with  $J = 20$  basis functions. In this example, we see that the approximation holds in the range of observed frequencies.

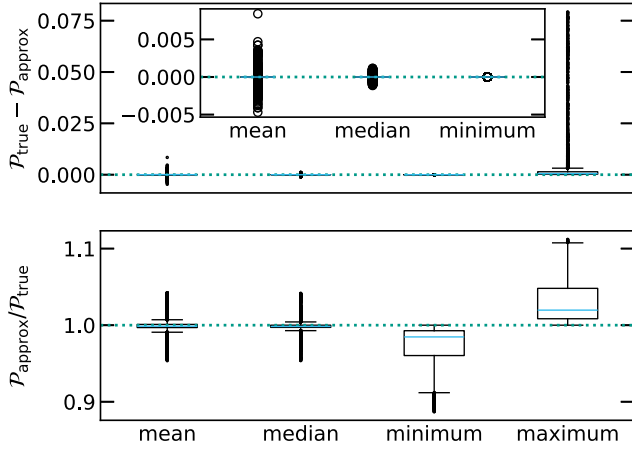
We also assess the approximation using other metrics such as the mean, median, maximum, and minimum values of the frequency-



**Figure 1.** True (solid line) and approximated (dashed-dot line) models with the basis functions (dashed lines) for a single bending (blue) and double bending (orange) power spectrum model.



**Figure 2.** Residuals (top) and ratios (bottom) between the intended power spectrum model and its approximation as a function of frequency. The median, the 68th and 95th percentiles are also shown. The minimum and maximum frequencies of the observed data are shown in black dashed lines.



**Figure 3.** Distributions of the mean, median, minimum and maximum of the frequency-averaged residuals (top) and ratios (bottom) between the true power spectrum model. The extreme values of each distribution are shown as whiskers.

averaged residuals and ratios. This is shown in Fig. 3, the residuals are mainly located around zero while the ratios are centred around one. A more detailed analysis of the choice of  $J$  is given in Appendix B2. The quality of the approximation depends on the model approximated, the basis function  $\psi$ , the number of basis functions  $J$  and the size of the frequency grid.

### 3.3 Other assumptions

We have assumed that the time series consists of Gaussian measurements. However, the distribution of flux in accreting black hole light curves appears to be log-normal (Uttley, McHardy & Vaughan 2005). This is thought to originate from multiplicative independent and identically distributed processes rather than additive processes. This can be observed as a linear relation between the mean flux and the root mean square amplitude of variability (Uttley & McHardy 2001; Uttley et al. 2005). While the origin and interpretation of such a relation is debated (Scargle 2020), we only intend to include a way to model the non-Gaussian distribution of fluxes. Assuming the distribution of fluxes is log-normal, taking the logarithm of the observed light curve  $x$  makes the data Gaussian as shown in equation (11). The additional parameter  $c$  accounts for a possible shift in the log-normal distribution.  $c$  could be intrinsic to the source variability or instrumental, for example background.

$$y = \ln(x - c) \quad \text{where} \quad 0 \leq c < \min x \quad (11)$$

Additionally, the transformation is propagated on the measurement uncertainties as shown in equation (12). The time series  $y$  can now be used in the GP regression with measurement uncertainties  $\sigma_y$ .

$$\sigma_y^2 = \frac{\sigma_x^2}{(x - c)^2} \quad (12)$$

Other transformations such as the Box–Cox transformation (Box & Cox 1964) could be considered to make the data normally distributed. Implicit in the assumption of a Gaussian time series with independent Gaussian measurement errors, we also assume that no outliers are present in the data.

If the time series is a combination of times series from multiple instruments, it is possible to include a cross-calibration factor. In Section 6, we apply the method to real data from two different

instruments and use a scale factor named  $\gamma$  on one of the time series.  $\gamma$  should be one when the cross-calibration is accurate.

### 3.4 Implementation

We implement the method in various packages, the code is available in the PYTHON library `stingray`<sup>2</sup> (Huppenkothen et al. 2019a, 2019b) using `tinygp`<sup>3</sup> for the GP regression with quasi-separable covariance functions using Just-in-time compilation with JAX (Bradbury et al. 2018).

A pure JULIA implementation is also available in `Pioran.jl`.<sup>4</sup> In Fig. 4, we compare the likelihood evaluation time between the implementations in PYTHON and JULIA, and also a Fast Fourier Transform-based implementation method (see Section 7.1).

The PYTHON implementation using `tinygp` appears to be an order of magnitude slower than the JULIA implementation but is faster than the direct method. This speed-up is also observed when comparing the C++ implementation of `celerite`<sup>5</sup> to `tinygp`.

## 4 BAYESIAN WORKFLOW

In this work, we adopt a Bayesian workflow (Gelman et al. 2020) for our modelling and to obtain estimates and credible intervals for the parameters of our models. The set of stages proposed in the workflow cover model building, inference, and checks before and after inference, all of this in a Bayesian framework. In Section 2.1.1, we defined the GP likelihood function, here we define the priors and methods we use in a typical workflow.

### 4.1 Bayes' rule

Bayes' rule in equation (13) gives the posterior probability density  $p(\theta|d)$  of parameters  $\theta$ , knowing that we observed data  $d$ . The likelihood function is denoted by  $p(d|\theta) = \mathcal{L}(\theta)$  and the prior probability density is given by  $p(\theta)$ .

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}. \quad (13)$$

The normalizing constant in equation (13) is called the marginal likelihood, Bayesian evidence or simply evidence and is given by:

$$Z = p(d) = \int p(d|\theta)p(\theta)d\theta. \quad (14)$$

The Bayesian evidence can be used to compare two models using the Bayes factor  $BF_{12} = Z_1/Z_2$ . When  $BF > 10$  one model can be favoured, with  $BF > 100$  the model is strongly supported (Jeffreys 1939; Bailer-Jones 2017). Sampling from the posterior probability distribution can be achieved with Markov chain Monte Carlo (MCMC) algorithms but to estimate the evidence at the same time, nested sampling algorithms are required (Skilling 2004).

### 4.2 Priors

Before performing any inference on a given time series, we first select a model. Then, we define a joint prior distribution which accounts for any information on the parameters of the model regardless of the data. Table 3 lists all the parameters used for the inference of a power

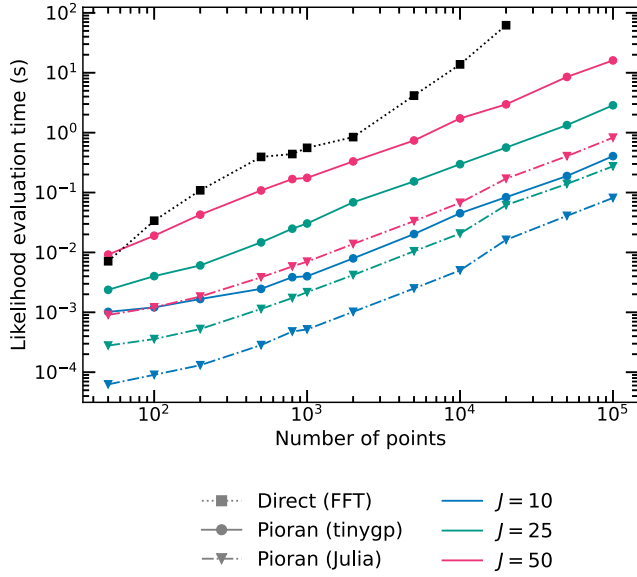
<sup>2</sup><https://docs.stingray.science/en/latest/>

<sup>3</sup><https://tinygp.readthedocs.io/en/stable/>

<sup>4</sup><https://github.com/mlfckir/Pioran.jl>

<sup>5</sup><https://celerite2.readthedocs.io>





**Figure 4.** Likelihood evaluation time in seconds for the approximation method and the direct method as a function of the number points and number of basis functions  $J$  for the approximation. The black squares show regression with direct method scaling as  $\mathcal{O}(N^3)$  (see Section 7.1). The circles and triangles respectively represent the PYTHON `tinygp` and JULIA `Pioran.jl` codes where colours encode the number of SHO basis functions.

spectrum model with two bends assuming a log-normal distributed time series.

#### 4.2.1 Power spectrum parameters

In the case of power-law power spectra, our prior beliefs are that the slopes must agree on a decreasing power spectrum. To ensure that the approximation holds, the power spectrum model must not be steeper than  $f^{-\alpha_{\max}}$  where  $\alpha_{\max}$  equals 4 in the case  $\psi_4$  and 6 for  $\psi_6$ . We also expect the power spectrum to be flat at low frequencies. As shown in Table 3, we use uniform priors on the slopes, where the slope  $\alpha_{i+1}$  depends on the previous slope  $\alpha_i$ . This is equivalent to defining first  $\alpha_1$  with a uniform prior  $p(\alpha_1)$ , and the prior distribution on  $\alpha_2$  is conditional on  $\alpha_1$  and then the prior distribution of  $\alpha_3$  is conditional on  $\alpha_2$  and so on. This can be expressed with  $p(\alpha_1)$  and  $p(\alpha_2|\alpha_1)$  in the case of a single-bending power law.

We require the low-frequency bend  $f_1$  to be in a sensible frequency range given by  $f_{\text{start}} = f_{\min}/S_{\text{low}}$  and  $f_{\text{stop}} = S_{\text{high}}f_{\max}$ , where  $f_{\min}$ ,  $f_{\max}$  denote the limiting frequencies of the time series

and  $S_{\text{low}}$ ,  $S_{\text{high}}$  are scale factors for the approximation. We use a log-uniform prior for the first bend, and similarly to the slopes, the prior on  $f_2$  depends on the  $f_1$ .

We do not model the amplitude of the power spectrum; instead, we include its total integral (from  $f = -\infty$  to  $f = +\infty$ ), which is the amplitude of the autocovariance function, as a parameter.

#### 4.2.2 Gaussian process parameters

To specify the prior on the mean of the time series, we extract a random subset of the total time series – between 1 and 3 per cent of the points – which we use to compute a sample mean  $\bar{x}$  and sample variance  $s^2$ . This subset is then discarded for the inference and the remainder of the analysis.

If the time series is assumed to be Gaussian then the prior for the mean can be set as a normal distribution centred on  $\bar{x}$  with variance  $\beta s^2$  where  $\beta$  is a scale factor. If we assume the time series to be log-normally distributed, then the values of  $\bar{x}$  and  $s^2$  are computed using the logarithm of the subset time series.

The amplitude of the autocovariance function is the variance of the process and should not be confused with the sample variance of the time series which is an estimator of the variance of the process. To specify its prior distribution in the context of accreting supermassive black holes, we use  $F_{\text{var}}$  the fractional root mean square variability amplitude defined in equation (15), where  $\sigma_{\text{err}}^2$  is the mean square error (Vaughan et al. 2003).

$$F_{\text{var}} = \sqrt{\frac{s^2 - \sigma_{\text{err}}^2}{\bar{x}^2}}. \quad (15)$$

When looking at previous analyses of time series from active galaxies (e.g. Markowitz et al. 2003a; Markowitz, Edelson & Vaughan 2003b),  $F_{\text{var}}$  is almost always less than one. We choose to model  $F_{\text{var}}$  with a log-normal distribution with parameters  $\mu_{F_{\text{var}}} = -1.5$  and  $\sigma_{F_{\text{var}}}^2 = 1/2$ . Given that the product of log-normal variables is also log-normal, then  $F_{\text{var}}^2$  also follows a log-normal distribution with mean  $2\mu_{F_{\text{var}}}$  and variance  $4\sigma_{F_{\text{var}}}^2$ . To define a prior for the variance of the underlying process we assume that  $F_{\text{var}}^2$  is proportional to the true variance of the process and thus assume that the distribution of  $F_{\text{var}}^2$  can be used as a prior knowledge on the variance. Therefore, we choose a log-normal distribution with parameters  $\mu = -3$  and  $\sigma^2 = 2$ .

We choose a Gamma(2, 0.5) prior on  $\nu$  as we expect its value to be close to 1 on average. The Gamma distribution is parametrized with the shape and scale parameters. In the case of a logarithmic transformation of the data (equation 11), the offset  $c$  in the logarithm has a log-uniform prior. In the case where  $c = 0$ , one can use a half-

**Table 3.** Parameters and priors for the modelling of a power spectrum with two bends. The first half lists the power spectrum parameters and the second half lists the Gaussian process parameters.

| Modelling                       | Parameter  | Description                                      | Prior distribution                                    | Unit          |
|---------------------------------|------------|--------------------------------------------------|-------------------------------------------------------|---------------|
| Power spectrum                  | $\alpha_1$ | Low-frequency slope                              | Uniform[0, 1.25]                                      | –             |
|                                 | $\alpha_2$ | High-frequency slope or intermediate slope       | Uniform[ $\alpha_1$ , $\alpha_{\max}$ ]               | –             |
|                                 | $\alpha_3$ | High-frequency slope                             | Uniform[ $\alpha_2$ , $\alpha_{\max}$ ]               | –             |
|                                 | $f_{b,1}$  | Low-frequency bend                               | Log-uniform[ $f_{\text{start}}$ , $f_{\text{stop}}$ ] | same as $1/t$ |
|                                 | $f_{b,2}$  | High-frequency bend                              | Log-uniform[ $f_{b,1}$ , $f_{\text{stop}}$ ]          | same as $1/t$ |
| Gaussian process or Time series | variance   | Variance of the process                          | Log-normal(–3, 2)                                     | same as $x^2$ |
|                                 | $\nu$      | Scaling on the measurements uncertainties        | Gamma(2, 0.5)                                         | –             |
|                                 | $\mu$      | Mean of the Gaussian time series                 | Normal( $\bar{x}$ , $\beta s^2$ )                     | same as $x$   |
|                                 | $c$        | Offset for a log-normal-distributed time series  | Log-uniform[ $10^{-6}$ , $0.99 \min(x)$ ]             | same as $x$   |
|                                 | $\gamma$   | Inter-calibration factor for the two time series | Log-normal(–0.1, 0.2)                                 | –             |



Cauchy prior as used in Section 6 when applying the method to Ark 564.

As mentioned in Section 3.2, it is important to check that adequate priors were chosen for the approximation to hold, this is a prior predictive check. One can also draw samples from the GP using these priors to ensure that the time series have realistic values for the mean and variance.

### 4.3 Inference

In this work, we use the two methods presented below to sample from the posterior probability distributions. As we use `Pioran.jl`, the JULIA implementation of the power spectral approximation, the sampling codes presented here interface well with `Pioran.jl`.

#### 4.3.1 Nested sampling

Nested sampling (NS) enables global parameter exploration and has a natural self-convergence criterion (Skilling 2004). Most NS algorithms only require the likelihood function and the joint prior distribution to be evaluated, i.e. the likelihood does not need to be differentiable. In this work, we use `ULTRANEST`<sup>6</sup> (Buchner 2021) a robust PYTHON nested sampling code which implements the MLFRIENDS algorithm (Buchner 2014, 2019). This implementation requires little tuning and can be parallelized on multiple cores with MPI. We use the default values of 400 live points and `frac_remain` = 0.01 to obtain samples from the posterior distribution and an estimate of the evidence. Convergence is reached until the target criteria are fulfilled, i.e. the effective sample size is at least 400 and the new samples contribute to less than 1 per cent to the evidence (`frac_remain` = 0.01). In practice, to call `ULTRANEST` in JULIA, we use the package `PyCall.jl` and `MPI.jl` (Byrne, Wilcox & Churavy 2021) to speed up the inference with parallelization.

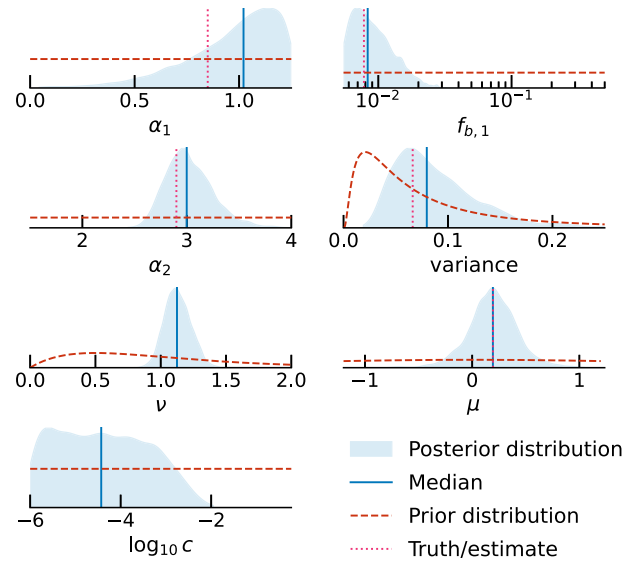
#### 4.3.2 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) (Betancourt 2017) with the No-U-Turn sampler (Hoffman & Gelman 2011) is an MCMC method that provides robust estimates of the posterior samples with little tuning. This method requires the posterior density to be differentiable which is the case in the GPs implementations of `tinygp` (PYTHON) and `Pioran.jl` (JULIA). We use the No-U-Turn sampler implemented in the JULIA package `AdvancedHMC.jl` (Xu et al. 2020) through the probabilistic programming library `Turing.jl` (Ge, Xu & Ghahramani 2018).

By default, we sample 2000 points from the posterior with a warm-up phase of 2500 iterations. We sample the posterior with 12 chains in parallel to use convergence diagnostics such as  $\hat{R}$  and the effective sample size (ESS) (Vehtari et al. 2021). For all parameters,  $\hat{R}$  should be close to unity and it is recommended that the ESS should be large enough (e.g. about  $\sim 400$ ). Furthermore, a visual inspection of the chains and the posterior samples can be performed as an additional check for convergence.

### 4.4 Diagnostics

To illustrate the diagnostics in this Section, we use the results of one of the simulations presented in Section 5. Once the posterior samples have been generated, we plot the distribution of the posterior samples



**Figure 5.** Distributions of the posterior samples for one simulated time series. The median of the distributions is shown by a blue vertical line while the true value is shown by a dotted magenta vertical lines. The prior distributions are shown with dashed red lines.

against the prior distribution as shown in Fig. 5. This allows checking that the priors are not too restrictive, we also compute the median of the distribution as shown by the vertical blue lines in the Fig. 5.

To check whether the inference yielded a ‘good fit’, we provide several visual diagnostics using samples from the posterior distribution. As an example, we use the results of one simulation and present these diagnostics in Fig. 6.

#### 4.4.1 In the time domain

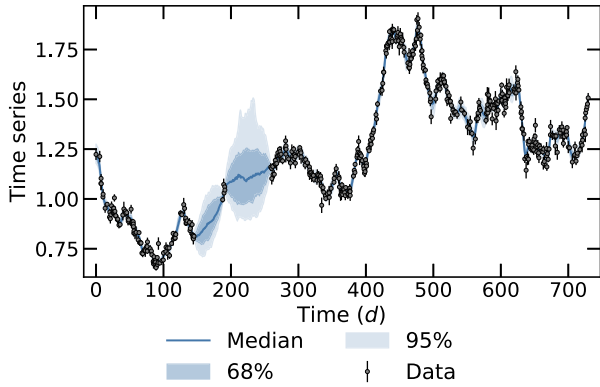
Given a model and parameter values  $\theta$ , we can predict the time series at any time by conditioning the GP on the observation. This assumes that we have a best-fitting value for the parameters of the models, for instance, Kelly et al. (2014) used a maximum a posteriori estimate.

Instead of using a single point estimate, here we use a subset of the posterior samples generated with MCMC or NS. For each set of parameter values, we condition the GP on the observed data and the error bars  $\sigma^2$  using equations (5) and (6) to obtain a conditioned GP. From this process, we then draw a realization. After iterating over the subset of posterior samples, we have a distribution of realizations, from which we compute the median and quantiles and study how accurate the GP is at predicting the data.

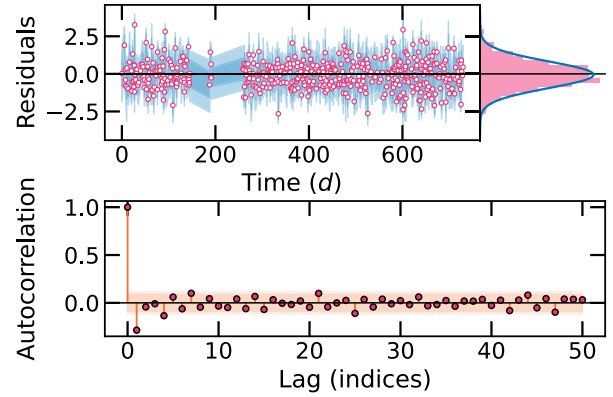
This is shown in Fig. 6(a) where the shaded areas show the quantiles of the predictive distribution. The noisy variations observed at the boundaries of prediction bands are due to the finite number of samples drawn, if we had drawn more samples these boundaries would appear smoother. It should be stressed that the covariance matrix of the conditioned GP may not retain the nice quasi-separable properties of the original GP, which makes sampling and drawing realizations from the conditioned GP very slow for large data sets.

Using the samples from the conditioned GP we also compute the standardized residuals defined as  $r = (y - y_{\text{sim}})/\sigma_y$ . If the observation is only contaminated by Gaussian white noise and if the model accurately reproduces the red noise time series, the residuals should be white, i.e. Gaussian distribution with zero-mean. In the upper panel of Fig. 6(b), we show the time series of the residuals and the

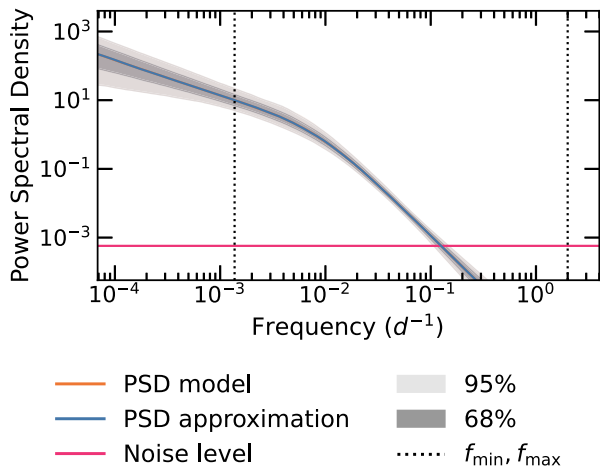
<sup>6</sup><https://johannesbuchner.github.io/UltraNest/>



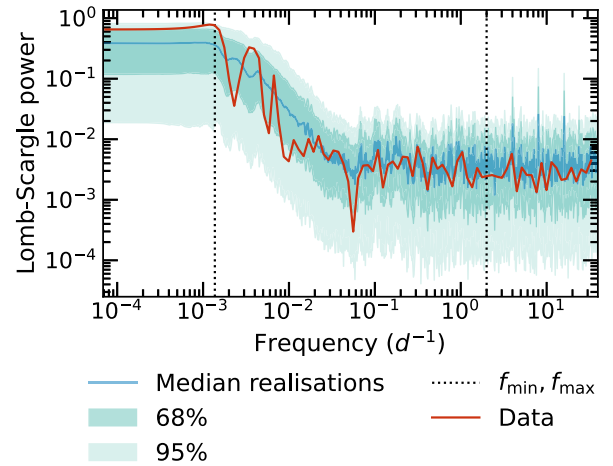
(a) Simulated predictive time series against the observed time series.



(b) Time series and distribution of the standardised residuals (top panel) and autocorrelation of the residuals (bottom panel).



(c) Posterior predictive power spectrum.



(d) Posterior predictive Lomb-Scargle periodogram from realisations of the conditioned Gaussian process.

**Figure 6.** Diagnostics post-inference using the posterior samples.

distribution of residuals. We observe that the residuals resemble a Gaussian white noise. Following Kelly et al. (2014), we plot in the lower panel the autocorrelation function (ACF) of the residuals, if the residuals are not equispaced in time then the time lags of the ACF are not proper time lags but lags between indexes. This may explain the deviation of the second value of the ACF from the expected distribution of white noise.

A possible way to check if non-equispaced residuals follow a white noise distribution could be to fit a first-order autoregressive process to the residuals. If the autoregressive coefficient is close to zero then residuals follow a white noise.

#### 4.4.2 In the Fourier domain

First, it is recommended to check that the power spectrum approximation in terms of  $J$  basis functions still holds with the posterior samples. This can be checked by comparing the power spectrum model, and the approximated power spectrum model using the posterior samples as shown in Fig. 6(c). In this case, we do not see any notable difference between the model in orange and the approximation in blue. If the approximated power spectrum model were to deviate from the intended model (bending power law), for

example if we discern the shape of individual basis functions in the approximated model then we advise increasing the number of basis functions and restarting the inference. As the normalization of the power spectrum is defined with the total variance, the noise level is given by  $P_{\text{noise}} = 2\nu\Delta t\sigma_{\text{err}}^2$  (see appendix A of Vaughan et al. 2003).

Diagnostics in the Fourier domain are critical to assess whether features were missed in the modelling. We suggest drawing realizations from the GP – using the same sampling pattern – with samples from the posterior distribution and computing the Lomb–Scargle periodogram (Lomb 1976; Scargle 1982). In Fig. 6(d), we compare the distribution of these periodograms to the Lomb–Scargle periodogram of the observation and find a good agreement. If we find some features missing in the periodogram of the realizations, we could include them in the modelling and compare the two models. We stress that this diagnostic based on the periodogram of irregular time series in the Fourier domain is challenging and may strongly depend on the sampling pattern. Other diagnostics in the Fourier domain could involve the Fourier transform of the discrete correlation function (Edelson & Krolik 1988). The purpose of these diagnostics is to ensure that the model is not completely wrong. For instance, these are not used to check whether the estimated bend(s) or slopes agree with the Lomb–Scargle periodogram which is distorted by the

sampling pattern and can be noisy. This can be seen in Fig. 6(d) where the periodogram does not show signs of a bend.

## 5 SIMULATION-BASED CALIBRATION

To validate the method presented here, we simulate several types of fake time series and try to recover the input parameters.

### 5.1 Simulating random time series

We simulate a time series  $x$  given a power spectrum model following the method of Timmer & Koenig (1995) (see also Ripley 1987). To include any effect of aliasing or leakage, the time series is simulated with a longer duration and a finer sampling than what is required, and we use only a subset of this simulated time series. In the limit of an infinite number of data points, this method is equivalent to drawing realizations from a GP.

To produce a log-normal distribution of values for  $y$ , the time series  $x$  is exponentiated, i.e.  $y_i = \exp x_i$ . For each simulated value, a measurement error is simulated as  $\epsilon_i = \beta \sqrt{y_i}$  where  $\beta$  is a real number drawn from the interval [1,4] where the probability is such that  $P(\beta = 1) = 0.99$  and  $P(\beta = 4) = 0.01$ . To simulate the observation process, the observed value  $y_i^{\text{obs}}$  is drawn from the Gaussian distribution:  $\text{Normal}(y_i, \epsilon_i^2)$ .

Finally, we generate random gaps in the time sampling. The number and duration of the gaps are drawn from uniform distributions with the condition that the duration of the gappy time series should be no less than 75 per cent of the input duration. In our simulations, we set the condition that in the final time series,  $F_{\text{var}}$  should not exceed 1.5. In Table 4, we present the two sets of simulations used in this work. We simulate time series with a single-bending power-law power spectrum for short and long observations. For the short observation, we also simulate time series with a steep power spectrum, i.e. with a high-frequency slope  $\alpha_2 \sim \text{Uniform}[3, 6]$ .

### 5.2 Priors and inference

To infer the power spectrum of the simulated time series, we use the NS and HMC codes presented in the previous section. NS is used for all data sets while HMC is only used on the short data set. For the data set of short time series, two types of priors for the slope  $\alpha_2$  are used. We use prior distributions identical to the input of simulations shown in Table 4, or with a prior on  $\alpha_2$  conditional on the value of  $\alpha_1$ , this can be written as  $\pi(\alpha_2|\alpha_1)$ .

**Table 4.** Summary of the simulation set-up with details about the time series and the input model.

| Data set              | Short                                                                                                        | Long  |
|-----------------------|--------------------------------------------------------------------------------------------------------------|-------|
| Duration              | 2 yr                                                                                                         | 16 yr |
| Minimum sampling      | 6 h                                                                                                          | 6 h   |
| Number of points      | 500                                                                                                          | 1000  |
| Number of simulations | 4000                                                                                                         | 4000  |
| Number of gaps        | 0–3                                                                                                          | 3–6   |
| Bending power law     | $\alpha_1 \sim \text{Uniform}[0, 1.25]$                                                                      |       |
| Parameters            | $f_1 \sim \text{Log-uniform}[4f_{\text{min}}, f_{\text{max}}/4]^a$<br>$\alpha_2 \sim \text{Uniform}[1.5, 4]$ |       |
| Steep model           | $\alpha_1$ and $f_1$ identical as above                                                                      | –     |
|                       | $\alpha_2 \sim \text{Uniform}[3, 6]$                                                                         | –     |

<sup>a</sup>The values of  $f_{\text{min}}$  and  $f_{\text{max}}$  depend on the duration and sampling of the observation.

Prior parameters for the time series or the GP are the same as shown in Table 3. We also account for the log-normal distribution of values with a logarithmic transformation as the time series were exponentiated. For all simulations, the prior on the mean is defined with a zero-mean normal distribution with a variance of 4. The prior distribution of the variance is given by a log-normal distribution with parameters  $\mu = s^2$  and  $\sigma^2 = 1$ , where  $s^2$  is the sample variance of the time series. In this case, the prior distribution of the variance is not exactly Bayesian as it depends on the data. In these simulations,  $\mu$ ,  $\nu$ ,  $c$  and the variance are considered as nuisance parameters, we are interested in checking if we can recover the parameters of the power spectrum model.

### 5.3 Validating the method

Having defined a modelling framework, we now wish to establish its veracity. Computing the median or mean of the posterior distribution and comparing it to the true value may not be adequate in a Bayesian framework as they are simply point estimates. We use simulation-based calibration (Cook, Gelman & Rubin 2006; Talts et al. 2018) to assess the quality of the method and ensure that the distribution of the posterior samples is consistent with the input distribution of the simulated data sets. This method can be summarized as follows: draw parameter samples from the prior distribution, simulate time series using the samples, collect posterior samples for all parameters and simulations, and then compute the rank statistic of the prior sample relative to the posterior sample. The rank statistic can be seen as the index of the prior sample if it were to be inserted in an ordered sample of posterior values. Talts et al. (2018) showed that the rank statistic is expected to follow a discrete uniform distribution if the samples from the posterior distribution match the true underlying posterior.

In Fig. 7, we present the graphical diagnostics introduced in Säilynoja, Bürkner & Vehtari (2021) to assess the method. They diagnose when the distribution of the rank statistic (top row) deviates from its expected uniform distribution using the empirical cumulative distribution function (ECDF) in the middle and bottom rows. The second and third columns show that  $f_{b,1}$  and  $\alpha_2$  follow their expected distribution. We find that  $\alpha_1$  may be overestimated, however, if we restrict the simulations to larger values of  $f_{b,1}$  this bias vanishes. This means that when the bend is close to the minimum observed frequency, the estimation of the low-frequency slope is difficult and can be biased. We find however that the posterior distributions are not over or underdispersed, meaning there is no overfitting. Inference with NS and HMC sampling algorithms agrees in most cases.

## 6 APPLICATION TO ARK 564

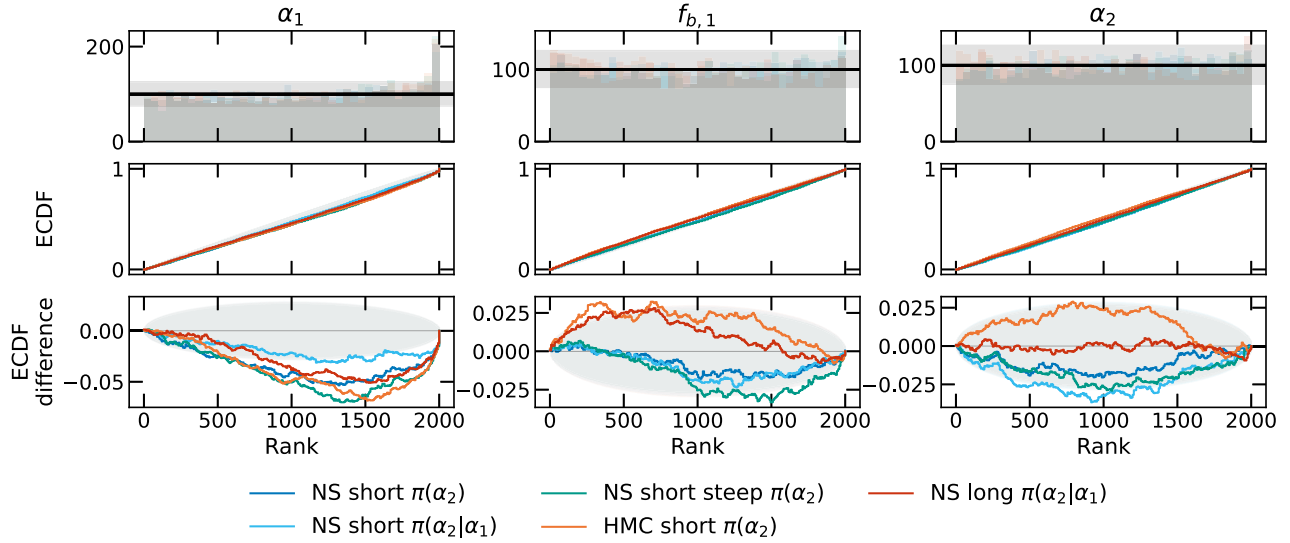
We demonstrate how this method can model the power spectrum of X-ray fluctuations in the active galaxy Ark 564. We use observations from the *Neil Gehrels Swift Observatory* (Gehrels et al. 2004) and *XMM-Newton* (Jansen et al. 2001) spanning nearly 22 yr.

### 6.1 Data reduction and calibration

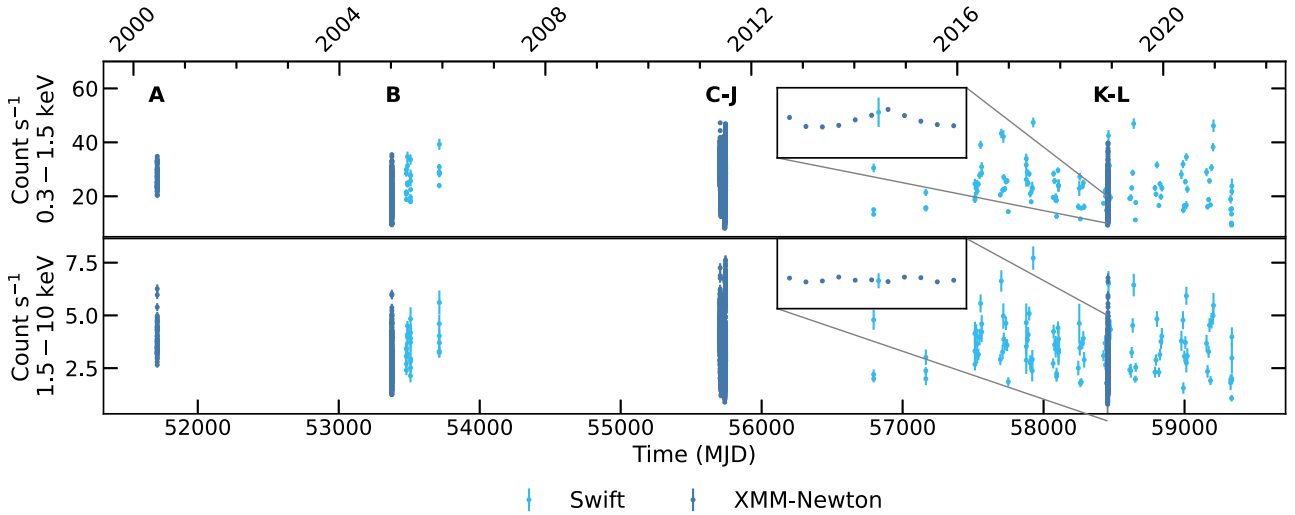
#### 6.1.1 Swift-XRT

We extract X-ray light curves from the *X-ray telescope* (Burrows et al. 2005) on board *Swift* with the online tool<sup>7</sup> described in Evans et al. (2007, 2009). The light curves are binned by snapshots, which means

<sup>7</sup>[https://www.swift.ac.uk/user\\_objects/](https://www.swift.ac.uk/user_objects/)



**Figure 7.** Simulation-based calibration. The top row shows the distribution of the rank statistic for the three parameters, and the horizontal black line shows the expected uniform distribution for the rank with a 99 per cent confidence band. The middle row shows the empirical cumulative distribution function (ECDF) and the bottom row shows the difference between the ECDF and the theoretical CDF. The band represents a 99 per cent confidence level of the theoretical CDF. NS and HMC, respectively correspond to inference with nested sampling and Hamiltonian Monte Carlo.  $\pi(\alpha_2|\alpha_1)$  indicates that the prior on  $\alpha_2$  is conditional on the value of  $\alpha_1$ . Short, short steep, and long correspond to the length of the simulations presented in Table 4.



**Figure 8.** Combined *Swift* and *XMM-Newton* light curves in the soft (top) and hard (bottom) bands. The insets show the simultaneous period between *XMM-Newton* and *Swift*. The letters refer to the *XMM-Newton* observations shown in Fig. 9 and presented in Table 5.

that we have one point per orbit of the spacecraft. The light curves are corrected for pile-up, bad columns on the CCD, and vignetting. We extract the light curves in the soft (0.3 – 1.5 keV) and hard (1.5 – 10 keV) energy bands. We only keep data obtained in the photon counting mode (PC). Fig. 8 shows the *Swift-XRT* light curve (light blue).

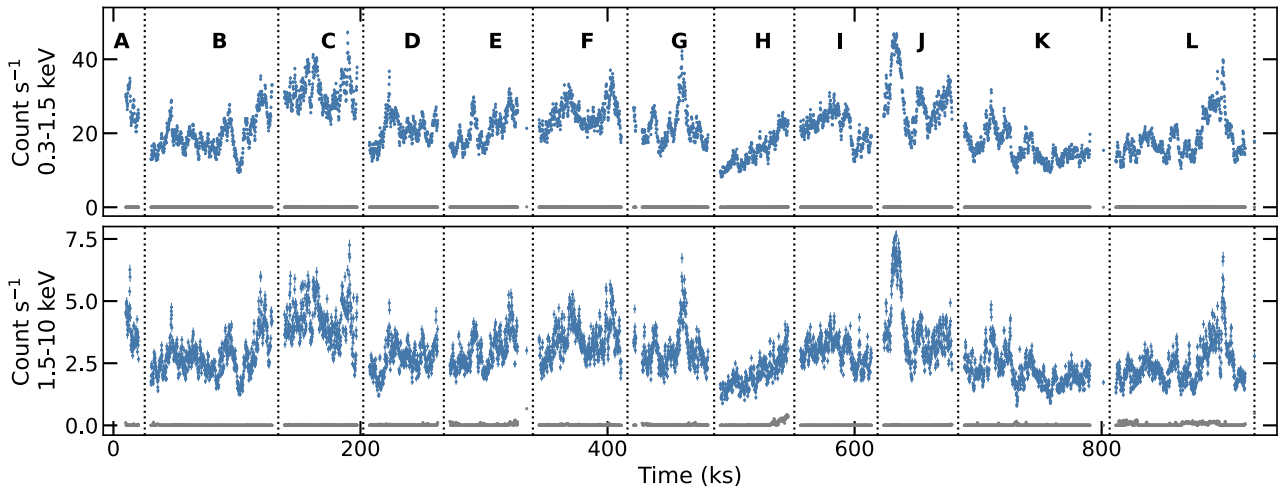
### 6.1.2 *XMM-Newton*

We select 12 *XMM-Newton* observations from 2000 to 2018 listed in Table 5. Data from the EPIC-pn camera (Strüder et al. 2001) were reduced using the Science Analysis System (SAS) 21.0 (Gabriel et al. 2004). All these observations were performed in small window mode, we selected the source in a circular region of radius 25 of

arcseconds. The background was selected in a source-free circular region of 50 arcsec radius. First, event lists for the source and background regions are extracted with `PATTERN<=4` to include single and double events. Then, counts are binned in light curves with a bin time of  $\Delta t = 150$  s. Finally, light curves are corrected for bad time intervals and loss of exposure.<sup>8</sup> No pile-up was noticeable in the observations. The *XMM-Newton* light curves in the soft and hard bands are shown in Fig. 9 (dark blue).

<sup>8</sup>Bins containing one or several bad time intervals have their exposure reduced, to correct for this effect the number of counts in the bin is rescaled proportionally to the lost exposure. Bins with zero exposure are removed.





**Figure 9.** Concatenated *XMM-Newton* light curves in the soft (top) and hard (bottom) energy bands. The background rate is shown in grey. The letters refer to the *XMM-Newton* observations presented in Table 5.

### 6.1.3 Inter-calibration

To combine the two light curves, we compare the count-rate for both instruments during a simultaneous observation in 2018 as shown in the inset of Fig. 8. A scaling factor  $\gamma$  is applied on the *Swift* data to match the *XMM-Newton* count-rate. The scaling factor is obtained by averaging the two *XMM-Newton* data points enclosing the *Swift* data. We find that the *Swift* count-rate must be multiplied by a factor 12.0 and 6.56 for the soft and hard bands, respectively.

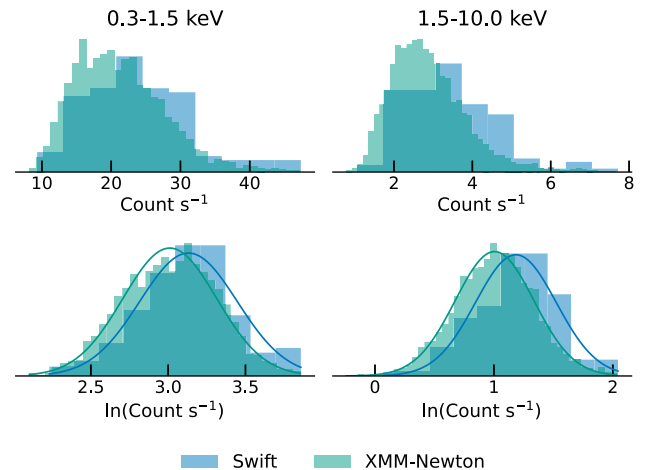
In Fig. 10, the distributions of the count-rates are shown, and the bottom panels show the distribution of the logarithm of these values. The distribution of the log of the count-rates appears to be well-modelled with a normal distribution. The distributions for both instruments overlap but do not perfectly match, an additional parameter will be added to account for the calibration.

## 6.2 Modelling

We infer the shape and parameters of the single and double-bending power-law model. As introduced in Section 3.2, we extend the low and high frequencies with the scale factors  $S_{\text{low}} = S_{\text{high}} = 20$ . We use  $J = 30$  basis functions. The priors on the power spectrum parameters

**Table 5.** *XMM-Newton* observations used in this work with the net exposure time and the number of points in the extracted light curve with  $\Delta t = 150$  s. The light curves of all epochs are plotted in Fig. 9.

| Epoch | Observation  | Date         | Net exposure (ks) | N   |
|-------|--------------|--------------|-------------------|-----|
| A     | 0006 810 101 | 2000 June 17 | 10.3              | 70  |
| B     | 0206 400 101 | 2005 Jan 05  | 98.6              | 654 |
| C     | 0670 130 201 | 2011 May 24  | 58.8              | 393 |
| D     | 0670 130 301 | 2011 May 30  | 55.2              | 369 |
| E     | 0670 130 401 | 2011 June 05 | 54.8              | 366 |
| F     | 0670 130 501 | 2011 June 11 | 66.6              | 445 |
| G     | 0670 130 601 | 2011 June 17 | 60.2              | 368 |
| H     | 0670 130 701 | 2011 June 25 | 54.9              | 367 |
| I     | 0670 130 801 | 2011 June 29 | 57.5              | 384 |
| J     | 0670 130 901 | 2011 July 01 | 55.2              | 368 |
| K     | 0830 540 101 | 2018 Dec 11  | 101.7             | 679 |
| L     | 0830 540 201 | 2018 Dec 03  | 104.7             | 695 |



**Figure 10.** Count-rate distribution of *XMM-Newton* and *Swift* light curves in the soft (left) and hard (right) energy bands (top panels). Distribution of the logarithm of the count-rates fitted with a normal distribution (bottom panels).

**Table 6.** Initial values for the Gaussian prior on the mean for the log-transformed soft and hard energy bands.

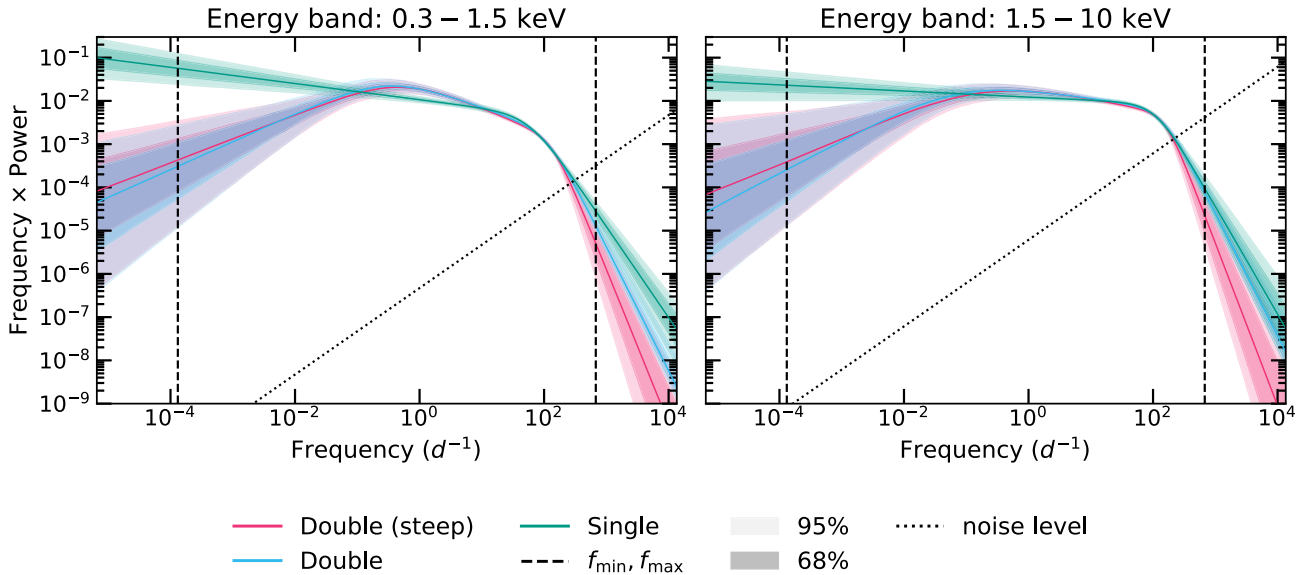
|                                             | 0.3 – 1.5 keV | 1.5 – 10 keV |
|---------------------------------------------|---------------|--------------|
| $\bar{x}$ (Count s <sup>-1</sup> )          | 3.01          | 1.01         |
| $s^2$ (Count <sup>2</sup> s <sup>-2</sup> ) | 0.0920        | 0.117        |

were presented in Table 3. We apply a logarithmic transformation to the data, as presented in Section 3.3. We tried three different distributions for the prior on the shift of the logarithm transformation. A log-uniform distribution, a Half-Cauchy distribution and a log-Half-Cauchy distribution, all yielded identical posterior distributions for all other parameters with similar Bayesian evidence. We present the results with the log-Half-Cauchy prior in Table 7.

To estimate the mean and variance of the process, we use the priors defined previously in Section 4.2.2. We extract 3 per cent of the original time series to compute the initial values for the prior on the mean, these values are presented in Table 6. We account for

**Table 7.** Results of the power spectrum estimation of Ark 564. Median values of the posterior distributions with the 16th and 84th percentiles. The evidence  $\ln Z$  and the logarithm of the Bayes factor comparing against the single-bending power-law model.

| Parameter                                                 | 0.3 – 1.5 keV             |                            |                            | 1.5 – 10 keV               |                            |                            |
|-----------------------------------------------------------|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
|                                                           | Single                    | Double                     | Double (steep)             | Single                     | Double                     | Double (steep)             |
| $\alpha_1$                                                | $1.19 \pm 0.04$           | $0.36^{+0.21}_{-0.30}$     | $0.44^{+0.18}_{-0.28}$     | $1.07 \pm 0.03$            | $0.25^{+0.33}_{-0.32}$     | $0.41^{+0.27}_{-0.37}$     |
| $f_{b,1} \text{ (d}^{-1}\text{)}$                         | $58.86^{+10.10}_{-10.12}$ | $0.31^{+0.31}_{-0.15}$     | $0.50^{+0.50}_{-0.25}$     | $110.37^{+12.45}_{-13.24}$ | $0.10^{+0.25}_{-0.06}$     | $0.25^{+1.13}_{-0.17}$     |
| $\alpha_2$                                                | $3.11^{+0.20}_{-0.17}$    | $1.53 \pm 0.06$            | $1.60 \pm 0.07$            | $3.49^{+0.28}_{-0.29}$     | $1.22^{+0.06}_{-0.05}$     | $1.29^{+0.10}_{-0.07}$     |
| $f_{b,2} \text{ (d}^{-1}\text{)}$                         | –                         | $116.87^{+11.92}_{-10.98}$ | $138.45^{+12.14}_{-13.34}$ | –                          | $139.09^{+13.78}_{-13.10}$ | $158.73^{+14.25}_{-12.53}$ |
| $\alpha_3$                                                | –                         | $3.87^{+0.10}_{-0.17}$     | $4.55^{+0.59}_{-0.47}$     | –                          | $3.84^{+0.12}_{-0.22}$     | $4.78^{+0.66}_{-0.61}$     |
| variance ( $\text{Count}^2 \text{ s}^{-2}$ )              | $0.58^{+0.30}_{-0.19}$    | $0.11^{+0.02}_{-0.01}$     | $0.11^{+0.02}_{-0.01}$     | $0.28^{+0.09}_{-0.07}$     | $0.12^{+0.02}_{-0.01}$     | $0.12^{+0.02}_{-0.01}$     |
| $\nu$                                                     | $0.41 \pm 0.14$           | $0.54^{+0.10}_{-0.11}$     | $0.75^{+0.15}_{-0.16}$     | $0.71 \pm 0.10$            | $0.70^{+0.09}_{-0.10}$     | $0.86^{+0.11}_{-0.12}$     |
| $\mu \text{ (Count s}^{-1}\text{)}$                       | $3.09 \pm 0.50$           | $3.07 \pm 0.07$            | $3.06 \pm 0.07$            | $1.04^{+0.29}_{-0.31}$     | $1.05 \pm 0.08$            | $1.05 \pm 0.08$            |
| $\ln c \text{ (Count s}^{-1}\text{)}$                     | $-23.13^{+9.26}_{-20.47}$ | $-23.34^{+9.40}_{-18.56}$  | $-23.22^{+9.10}_{-19.26}$  | $-23.92^{+9.19}_{-19.33}$  | $-23.39^{+8.71}_{-19.41}$  | $-23.64^{+9.20}_{-19.58}$  |
| $\gamma$                                                  | $0.94 \pm 0.07$           | $0.93 \pm 0.06$            | $0.92^{+0.07}_{-0.06}$     | $0.83^{+0.10}_{-0.09}$     | $0.86 \pm 0.07$            | $0.85 \pm 0.07$            |
| $\ln Z$                                                   | 7639.12                   | 7663.33                    | 7665.48                    | 3858.01                    | 3865.96                    | 3869.99                    |
| Log Bayes factor $\ln BF = \ln Z - \ln Z_{\text{single}}$ | –                         | 24.21                      | 26.36                      | –                          | 7.95                       | 11.98                      |

**Figure 11.** Posterior predictive power spectra for the soft (left) and hard band (right). The single-bending power law is plotted in green, the double-bending power law is shown in blue, and the double-bending power law with steep basis functions is shown in pink. The dark shaded areas represent the 68 per cent confidence while the light areas represent the 95 per cent confidence.

the log-normal distribution of count rates using the transformation presented in Section 3.3. As we find that the high-frequency slope  $\alpha_3$  may be steeper than 4, we also use the steep basis function  $\psi_6$  for the approximation.

## 6.3 Results and discussion

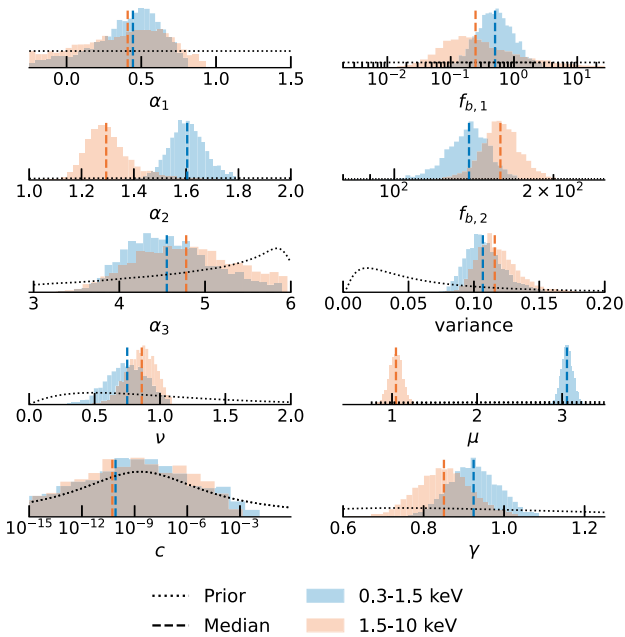
### 6.3.1 Results

The posterior samples for all the models and energy bands are shown in Figs C1 and C2 of Appendix C. Posterior medians and uncertainties from the 16th and 84th percentiles are given in Table 7. The last row shows the logarithm of the Bayes factor against the single-bending power-law model. We find very strong evidence for

two bends in the power spectrum of Ark 564 as the Bayes factors are all very high ( $BF \gg 10^2$ ).

The *XMM-Newton*/*Swift* inter-calibration factor  $\gamma$  is close to unity which shows that empirical inter-calibration using the simultaneous segment is correct. We find that the factor  $\nu$  for uncertainties on measurements is less than unity. This signifies that the uncertainties are overestimated. The noise level in the power spectrum defined as  $2\Delta t \nu \sigma_{\text{err}}^2$  is then decreased.

The posterior predictive power spectra for the three models are shown in Fig. 11 for the soft and hard bands. The two bends are distinguishable and we observe the difference between the approximations using  $\psi_4$  and  $\psi_6$ . Fig. 12 shows the posterior distributions for all the parameters of the double-bending model approximated using  $\psi_6$ , the prior distributions are shown with dotted lines. The median value on  $c$  the shift in the logarithm transformation is very



**Figure 12.** Distribution of the posterior samples for the double-bending power law approximated with  $\psi_6$ , in the soft band (blue) and hard band (orange). The medians are shown with vertical dashed–dotted lines. The priors are shown in black dashed lines.

small, this is due to the posterior distribution being nearly equal to the prior distribution as observed in the figure.

### 6.3.2 Comparison with previous works

Using *RXTE*, *ASCA*, and *XMM–Newton* observations McHardy et al. (2007) found that the periodogram of Ark 564 was well modelled with either a double-bending power-law model or with two Lorentzians. The energy bands are slightly different from ours: 0.6 – 2 keV for the soft band and 2 – 8.8 keV for the hard band. The low-frequency slope was fixed to zero in their modelling.

We find that all our estimates of the bends and slopes strongly agree with the results previously obtained. This was done using nearly independent data – only the 2005 *XMM–Newton* observation is used in both works. Fig. 12 shows that the low-frequency slope  $\alpha_1$  is consistent with zero and that the intermediate slope is steeper for the soft band. The high-frequency slope is steeper than 3.5 with a bend corresponding to a time-scale of about 10 min. It is difficult to distinguish differences in the posterior distributions of  $f_1$ ,  $f_2$ , and  $\alpha_3$  between the soft and hard bands.

## 7 DISCUSSION

Previous work has shown that the variability of AGN (and also often other accreting compact objects such as X-ray binaries and cataclysmic variables) is dominated by broad-band noise with a power spectrum that can be described reasonably well with power laws connected by smooth bends. The bend frequencies represent characteristic time-scales of the variable system and must be related to the size, geometry, and physics of the accretion flow and radiative mechanisms. Here, we have developed a new method, *PIORAN*, specifically for fitting this form of power spectral model and parameter estimates, using a fast GP approach which avoids the biases

and lack of calculable likelihood function associated with previous Fourier-based forward-modelling approaches. We now compare our method with other GP approaches for fitting power spectra without the constraints present in for example DRW approach, before discussing areas where *PIORAN* might be improved.

### 7.1 Standard Gaussian process methods

As introduced in Section 2.1, GPs allow modelling of power spectral densities, where the functional form of the covariance constrains the shape of the power spectrum. Miller et al. (2010) and Zoghbi et al. (2013) proposed to specify a power spectrum model, apply the discrete Fourier transform to compute numerically the autocovariance function and then use it in the GP regression.

This method suffers from two drawbacks, first, it cannot be applied to large data sets ( $N \gtrsim 10^3$ ) as it uses the standard GP approach which is computationally expensive (see Section 2.1.1). Secondly, with some tests, we find that when the grid of frequencies is large ( $\gtrsim 10^6$ ), the fast Fourier transform becomes a bottleneck. This can arise when a time series contains values separated by long time-scales and short time-scales, which can be the case for long-term light curves from *Swift* and *RXTE*.

In comparison, our method makes use of bending power-law power spectral model built using computationally efficient power law-like basis functions. Thanks to the *celerite* algorithm (Foreman-Mackey et al. 2017), our method is faster by at least an order of magnitude (see Fig. 4) and can be applied to much more data points – up to  $N \simeq 50,000$  in practice.

In Kelly, Sobolewska & Siemiginowska (2011), the power spectrum is modelled with a weighted sum of Lorentzian basis functions,  $\psi_2$ . The basis functions are also geometrically spaced in frequency, the important difference with our method relies on the weights. In Kelly et al. (2011), the weights allow modelling of a bending power-law power spectrum described as follows: flat at low frequencies, bends to  $f^{-\alpha}$  at an intermediate frequency, and then bends to  $f^{-2}$  at high frequencies. This weighting scheme allows estimation of the intermediate slope  $\alpha$  and the two bend frequencies. Here, we allow for arbitrary slopes at low, intermediate, and high frequencies as long as they agree on a decreasing power spectrum. Furthermore, our modelled power spectra can be steeper as the basis functions  $\psi_4$  and  $\psi_6$  are steeper than  $\psi_2$ .

CARMA processes of order  $(p, q)$  introduced in Section 2.2.2 allow flexible modelling of the broad-band variability in the time domain using a SDE. However, the interpretation of the coefficients of the SDE can be difficult and the choice of  $p$  and  $q$  is non-trivial as the models are not nested. Some simpler models that can be expressed with a few parameters (e.g. bending power law) may need very high-order CARMA representation, with complex relations between coefficients. In practice, this modelling makes the parametrization of the priors very difficult. Sampling from the posterior can be challenging as it will be multimodal when  $p$  and  $q$  are high.

In comparison, our method allows a more constrained shape for the broad-band power spectrum (bending power laws) and a simpler parametrization using only the parameters of the bending power-law model. CARMA processes can model narrow features in the power spectrum, with our method it could be possible to add such features manually after the bending power-law model is approximated.

### 7.2 PIORAN

Like all GP regression methods, *PIORAN* assumes Gaussian data and sampled values in the time series. While this method can be applied

to most AGN time series, here, we describe ideas on how one could lift some of these assumptions.

### 7.2.1 Stationarity

We assumed the underlying process generating the time series to be stationary. In practice, this may not be true for accreting systems. For instance, it is well-known that the energy spectrum of X-ray binaries can drastically transition from soft to hard state (McClintock & Remillard 2006). The physical processes dominant for the emission of photons are not the same between these states. For possibly different reasons, a similar behaviour is observed in the class of *Changing-look* AGN where the AGN transitioned from one class to another (Ricci & Trakhtenbrot 2023).

A way to study non-stationary signals could be to infer the power spectrum of contiguous segments of the time series and check whether they agree on the same model. However, choosing the number and duration of the segments might be challenging without any spectral information. A more general approach could be to model the non-stationarity with deep-state Gaussian processes (Zhao, Emzir & Särkkä 2020).

### 7.2.2 Gaussianity and sampling

In this work, we assume log-normal fluxes and apply a logarithmic transformation to the data to make the data Gaussian. As we use Gaussian processes, outliers are expected to be very rare, this means that tidal disruption events (TDEs), outbursts or flaring events cannot be well-modelled. A simple solution could be to filter these events from the time series by visual inspection of the posterior predictive time series after inference. A distribution with a broader tail might be more robust to outliers, for instance a Student- $t$  process (e.g. Tracey & Wolpert 2018).

Another process to consider is the photon-counting nature of X-ray counts, which can be modelled with a Poisson distribution. Assuming the background is negligible, the time series is then Poisson-distributed. Using a hierarchical Bayesian modelling one could account for Poisson data with the approach presented in this paper. However, this may be computationally expensive as each value of the time series will become a parameter of the hierarchical model. In practice, one can use approximate Bayesian methods such as variational inference to infer credible values for the process.

One should note that even periodogram-based methods also assume sampled values rather than binned values. It is possible to account for binned values in the GP framework with uncertain inputs. However, similarly to Poisson distributions, this requires a hierarchical Bayesian model which is computationally expensive.

### 7.2.3 Uncorrelated noise

In our modelling of the time series, we assumed the measurement variances to be independent and identically distributed. This is modelled by adding the term  $\sigma^2 I$  to the diagonal of the covariance matrix. In the case of time series from binned X-ray counts, the errors are uncorrelated but other time series may not have this property. For instance, it is critical to model the observation noise for exoplanet detection. Correlated noise can be modelled with a more complex model, see for instance the S+LEAF model in Delisle, Hara & Ségransan (2020) which allows banded noise models.

### 7.2.4 Quasi-periodic oscillations

The models used in the present paper do not include QPOs, but can easily be extended to allow for QPOs or other power spectral features. The bending power-law power spectrum would still be approximated using basis functions and one or more additional basis functions would model narrow QPO features. For strictly periodic variability, one could add a sinusoidal autocovariance function (which corresponds to a Dirac delta in the Fourier domain) or a periodic mean function to the GP model in the time domain. These additional features need to be added after approximating the continuum power spectrum model to avoid breaking the approximation. Then, Bayes factors (e.g. Anilkumar 2024), or posterior predictive  $p$ -values (Protassov et al. 2002; Gúrpide & Middleton 2025) can be used for model comparison (aperiodic noise with and without a QPO or strict period).

## 8 CONCLUSIONS

In this work, we presented a novel method to infer the parameters of bending power-law power spectra models for arbitrarily sampled time series. This method relies on GP regression which has no assumption about the time sampling pattern and avoids leakage and aliasing biases of the Fourier methods. It makes full use of data, time, value, and error and allows for heteroscedastic data. The method approximates the power spectrum model as a sum of  $J$  power law like basis functions. With this approximation, the power spectra models cannot be steeper than  $f^{-6}$ , but we believe it should be steep enough for time series of accreting compact objects. Our method relies on the fast and reliable algorithm presented in Foreman-Mackey et al. (2017) which enables computation of the log-likelihood with a linear scaling with respect to the number of data points in the time series. The workflow can be summarized as follows:

- (i) Select a bending power-law model with one or two bends, see Section 3.1.
- (ii) Define the number of basis functions ( $J$ ), the minimum and maximum frequencies for the approximation, see Section 3.2.
- (iii) If the time series data are not Gaussian, apply a transformation to make the data Gaussian, here we use a logarithmic transformation see Section 3.3.
- (iv) Choose suitable priors on the parameters of the power spectrum and time series, see Section 4.2.
- (v) Check the approximation by drawing samples from the prior and comparing with the selected model, see Section 3.2.1. If the number of basis functions is insufficient or if the priors are inadequate go back to (ii) and (iv).
- (vi) Sample the posterior distribution using HMC or NS methods, see Section 4.3.
- (vii) Perform inference and check for convergence Section 4.3
- (viii) Assess quality of the posterior samples with respect to the priors. See first paragraph of Section 4.4.
- (ix) Check for misspecification of the model using the diagnostics presented in Section 4.4. If the model is incomplete go back to (i) to change the model.
- (x) (Optional) Simulate time series from the posterior samples and use simulation-based calibration to ensure correctness of the workflow as presented in Section 5.

This workflow is fully Bayesian and applies to time series with arbitrary time sampling. Through extensive simulations, we checked that our method recovers the bending frequency and high-frequency slope unbiasedly with credible posterior distributions.



The low-frequency slope can be overestimated if the bending frequency is close to the minimum observed frequency of the time series.

Finally, we apply this method to long-term X-ray light curves of *XMM–Newton* and *Swift* observations of Ark 564. In agreement with McHardy et al. (2007), we find that the power spectrum is consistent with a double-bending power-law model. Our results strongly agree with previous analyses using observations from *RXTE*, *ASCA*, and *XMM–Newton*.

In a forthcoming paper, we will estimate the long-term X-ray power spectrum of a sample of unobscured active galaxies. The method presented here could have applications beyond X-ray astronomy, it could be used for future surveys such as Vera Rubin/LSST (Ivezić et al. 2019). Finally, this method could be extended to estimate delays and power spectra from multivariate time series as in Zu et al. (2013) and Wilkins (2019).

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their comments which improved the clarity of the text. ML acknowledges support from STFC studentships. ML thanks fruitful discussions about Fourier transforms and periodograms with Hervé Carfanten. ML thanks the University of Amsterdam and SRON for the warm welcome during the 4-week visit where part of this work was developed. The research leading to these results has received funding from the European Union’s Horizon 2020 Programme under the AHEAD2020 project (grant agreement no. 871158). This research used the ALICE High-Performance Computing Facility at the University of Leicester. This work has made use of observations obtained with *XMM–Newton*, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA. This work also made use of observations and data supplied by UK *Swift* Science Data Centre and has made use of data and/or software provided by the High Energy Astrophysics Science Archive Research Center (HEASARC), which is a service of the Astrophysics Science Division at NASA/GSFC.

*JULIA* packages: AdvancedHMC.jl (Xu et al. 2020), MPI.jl (Byrne et al. 2021), Julia (Bezanson et al. 2017), and Turing.jl (Ge et al. 2018).

*PYTHON* libraries: ARVIZ (Kumar et al. 2019), CORNER (Foreman-Mackey 2016), JAX (Bradbury et al. 2018), MATPLOTLIB (Hunter 2007), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), and ULTRANEST (Buchner 2021).

## DATA AVAILABILITY

The data used in this paper are publicly available to access and download from the *XMM–Newton* Science Archive and the UK Swift Science Data Centre. Final data products from this study can be provided on reasonable request to the corresponding author.

## REFERENCES

- Aigrain S., Foreman-Mackey D., 2023, *ARA&A*, 61, 329
- Ambikasaran S., 2015, *Numer. Linear Algebr. Appl.*, 22, 1102
- Anilkumar V., 2024, Master’s thesis, Anton Pannekoek Institute for Astronomy, Univ. Amsterdam
- Arévalo P., Uttley P., 2006, *MNRAS*, 367, 801
- Bachetti M., Huppenkothen D., 2022, preprint (arXiv:2209.07954)
- Bailer-Jones C. A. L., 2017, *Practical Bayesian Inference*. Cambridge Univ. Press, Cambridge
- Belcher J., Hampton J. S., Wilson G. T., 1994, *J. R. Stat. Soc. Ser. B Method.*, 56, 141
- Betancourt M., 2017, preprint (arXiv:1701.02434)
- Bezanson J., Edelman A., Karpinski S., Shah V. B., 2017, *SIAM Review*, 59, 65
- Box G. E. P., Cox D. R., 1964, *J. R. Stat. Soc. Ser. B Method.*, 26, 211
- Bradbury J. et al., 2018, JAX: composable transformations of Python+NumPy programs. <http://github.com/google/jax>, Accessed on 18/04/2025
- Brockwell P. J., Davis R. A., 2016, *Introduction to Time Series and Forecasting*, Springer Texts in Statistics. Springer International Publishing, Berlin
- Buchner J., 2014, *Stat. Comput.*, 26, 383
- Buchner J., 2019, *PASP*, 131, 108005
- Buchner J., 2021, *Stat. Surv.*, 17, 169
- Burrows D. N. et al., 2005, *Space Sci. Rev.*, 120, 165
- Byrne S., Wilcox L. C., Churavy V., 2021, *JCON*, 1, 68
- Chatfield C., 2004, *The Analysis of Time Series: An Introduction*, 6th edn, CRC Press, Florida
- Cook S. R., Gelman A., Rubin D. B., 2006, *J. Comput. Graph. Stat.*, 15, 675
- Delisle J. B., Hara N., Ségransan D., 2020, *A&A*, 638, A95
- Done C., Madejski G. M., Mushotzky R. F., Turner T. J., Koyama K., Kunieda H., 1992, *ApJ*, 400, 138
- Edelson R. A., Krolik J. H., 1988, *ApJ*, 333, 646
- Edelson R., Nandra K., 1999, *ApJ*, 514, 682
- Edelson R., Vaughan S., Malkan M., Kelly B. C., Smith K. L., Boyd P. T., Mushotzky R., 2014, *ApJ*, 795, 2
- Espinoza N., Kossakowski D., Brahm R., 2019, *MNRAS*, 490, 2262
- Evans P. A. et al., 2007, *A&A*, 469, 379
- Evans P. A. et al., 2009, *MNRAS*, 397, 1177
- Event Horizon Telescope Collaboration, 2019, *ApJ*, 875, L1
- Event Horizon Telescope Collaboration, 2022, *ApJ*, 930, L12
- Fabian A. C., 2012, *ARA&A*, 50, 455
- Foreman-Mackey D., 2016, *J. Open Source Softw.*, 1, 24
- Foreman-Mackey D., Agol E., Ambikasaran S., Angus R., 2017, *AJ*, 154, 220
- Gabriel C. et al., 2004, in Ochsenein F., Allen M. G., Egret D., eds, *ASP Conf. Ser.*, Vol. 314, *Astronomical Data Analysis Software and Systems (ADASS) XIII*. Astron. Soc. Pac., San Francisco, p. 759
- Ge H., Xu K., Ghahramani Z., 2018, in *International Conference on Artificial Intelligence and Statistics*. PMLR, p. 1682
- Gehrels N. et al., 2004, *ApJ*, 611, 1005
- Gelman A. et al., 2020, preprint (arXiv:2011.01808)
- González-Martín O., Vaughan S., 2012, *A&A*, 544, A80
- Gürpide A., Middleton M., 2025, *MNRAS*, 537, 3210
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Hoffman M. D., Gelman A., 2011, preprint (arXiv:1111.4246)
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Huppenkothen D. et al., 2019a, *J. Open Source Softw.*, 4, 1393
- Huppenkothen D. et al., 2019b, *ApJ*, 881, 39
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jansen F. et al., 2001, *A&A*, 365, L1
- Jeffreys H., 1939, *Theory of Probability*. Oxford Univ. Press, Oxford
- Jones R. H., 1981, in Findley D. F., ed., *Applied Time Series Analysis II*. Academic Press, p. 651
- Jones R. H., Ackerson L. M., 1990, *Biometrika*, 77, 721
- Kelly B. C., Bechtold J., Siemiginowska A., 2009, *ApJ*, 698, 895
- Kelly B. C., Sobolewska M., Siemiginowska A., 2011, *ApJ*, 730, 52
- Kelly B. C., Becker A. C., Sobolewska M., Siemiginowska A., Uttley P., 2014, *ApJ*, 788, 33
- Kozłowski S., 2016, *MNRAS*, 459, 2787
- Kumar R., Carroll C., Hartikainen A., Martin O., 2019, *J. Open Source Softw.*, 4, 1143
- Lomb N. R., 1976, *Ap&SS*, 39, 447
- Markowitz A., 2010, *ApJ*, 724, 26
- Markowitz A. et al., 2003a, *ApJ*, 593, 96
- Markowitz A., Edelson R., Vaughan S., 2003b, *ApJ*, 598, 935

Marshall N., Warwick R. S., Pounds K. A., 1981, *MNRAS*, 194, 987

McClintock J. E., Remillard R. A., 2006, Compact stellar X-ray sources. Cambridge Univ. Press, Cambridge, p. 157

McHardy I. M., Papadakis I. E., Uttley P., Page M. J., Mason K. O., 2004, *MNRAS*, 348, 783

McHardy I. M., Koerding E., Knigge C., Uttley P., Fender R. P., 2006, *Nature*, 444, 730

McHardy I. M., Arévalo P., Uttley P., Papadakis I. E., Summons D. P., Brinkmann W., Page M. J., 2007, *MNRAS*, 382, 985

Miller L., Turner T. J., Reeves J. N., Lobban A., Kraemer S. B., Crenshaw D. M., 2010, *MNRAS*, 403, 196

Mueller M., Madejski G., 2009, *ApJ*, 700, 243

Padovani P. et al., 2017, *A&AR*, 25, 2

Papadakis I. E., Lawrence A., 1993, *MNRAS*, 261, 612

Press W. H., 1978, Comments on Astrophysics, 7, 103

Priestley M. B., 1981, Spectral Analysis and Time Series. Academic Press, London

Protassov R., van Dyk D. A., Connors A., Kashyap V. L., Siemiginowska A., 2002, *ApJ*, 571, 545

Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning. The MIT Press, Cambridge, Massachusetts, p. 272

Rees M. J., 1984, *ARA&A*, 22, 471

Ricci C., Trakhtenbrot B., 2023, *Nat. Astron.*, 7, 1282

Ripley B. D., 1987, Stochastic simulation. John Wiley and Sons, Inc., New York

Säilynoja T., Bürkner P.-C., Vehtari A., 2021, preprint (arXiv:2103.10522)

Scargle J. D., 1981, *ApJS*, 45, 1

Scargle J. D., 1982, *ApJ*, 263, 835

Scargle J. D., 2020, *ApJ*, 895, 90

Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Ser., Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 395

Strüder L. et al., 2001, *A&A*, 365, L18

Summons D. P., Arévalo P., McHardy I. M., Uttley P., Bhaskar A., 2007, *MNRAS*, 378, 649

Talts S., Betancourt M., Simpson D., Vehtari A., Gelman A., 2018, preprint (arXiv:1804.06788)

Timmer J., Koenig M., 1995, *A&A*, 300, 707

Tracey B. D., Wolpert D. H., 2018, preprint (arXiv:1801.06147)

Uttley P., Malzac J., 2024, *MNRAS*, 536, 3284

Uttley P., McHardy I. M., 2001, *MNRAS*, 323, L26

Uttley P., McHardy I. M., 2005, *MNRAS*, 363, 586

Uttley P., McHardy I. M., Papadakis I. E., 2002, *MNRAS*, 332, 231

Uttley P., McHardy I. M., Vaughan S., 2005, *MNRAS*, 359, 345

van der Klis M., 1989, in Ögelman H., van den Heuvel E. P. J., eds, NATO Advanced Study Institute (ASI) Series C Vol. 262, Timing Neutron Stars. Kluwer Academic, New York, p. 27

Vaughan S., 2010, *MNRAS*, 402, 307

Vaughan S., Edelson R., Warwick R. S., Uttley P., 2003, *MNRAS*, 345, 1271

Vaughan S., Uttley P., Markowitz A. G., Huppenkothen D., Middleton M. J., Alston W. N., Scargle J. D., Farr W. M., 2016, *MNRAS*, 461, 3145

Vehtari A., Gelman A., Simpson D., Carpenter B., Bürkner P.-C., 2021, *Bayesian Anal.*, 16

Virtanen P. et al., 2020, *Nat. Methods*, 17, 261

Wilkins D. R., 2019, *MNRAS*, 489, 1957

Xu K., Ge H., Tebbutt W., Tarek M., Trapp M., Ghahramani Z., 2020, in Symposium on Advances in Approximate Bayesian Inference. PMLR, p. 667

Zhang H., Yan D., Zhang P., Yang S., Zhang L., 2021, *ApJ*, 919, 58

Zhang H., Yang S., Dai B., 2023, *ApJ*, 946, 52

Zhao Z., Emzir M., Särkkä S., 2020, preprint (arXiv:2008.04733)

Zoghbi A., Reynolds C., Cackett E. M., 2013, *ApJ*, 777, 24

Zu Y., Kochanek C. S., Kozłowski S., Udalski A., 2013, *ApJ*, 765, 106

## APPENDIX A: FOURIER TRANSFORMS

The basis function  $\psi_6(f)$  can be expressed as equation (A1), using partial fraction decomposition. The first term is the DRW basis function  $\psi_2(f)$  given in Table 2 with a known Fourier transform  $\phi_2(\tau)$ .

$$\psi_6(f) = \frac{1}{1+f^6} = \frac{1}{3} \left[ \frac{1}{1+f^2} + \frac{2-f^2}{1-f^2+f^4} \right] \quad (\text{A1})$$

The Fourier transform of the second term can be computed using Cauchy's residue theorem. The Fourier transform is given by equation (A2).

$$\mathcal{F} \left[ \frac{2-f^2}{1-f^2+f^4} \right] (\tau) = \pi e^{-\pi|\tau|} \left( \cos(\sqrt{3}\pi|\tau|) + \sqrt{3} \sin(\sqrt{3}\pi|\tau|) \right) \quad (\text{A2})$$

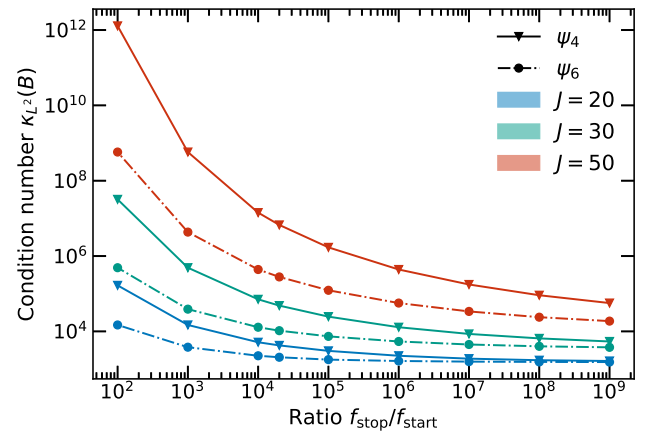
Thus  $\phi_6(\tau)$  is given by equation (A3).

$$\phi_6(\tau) = \frac{\pi}{3} \left[ e^{-2\pi|\tau|} + e^{-\pi|\tau|} \left( \cos(\sqrt{3}\pi|\tau|) + \sqrt{3} \sin(\sqrt{3}\pi|\tau|) \right) \right] \quad (\text{A3})$$

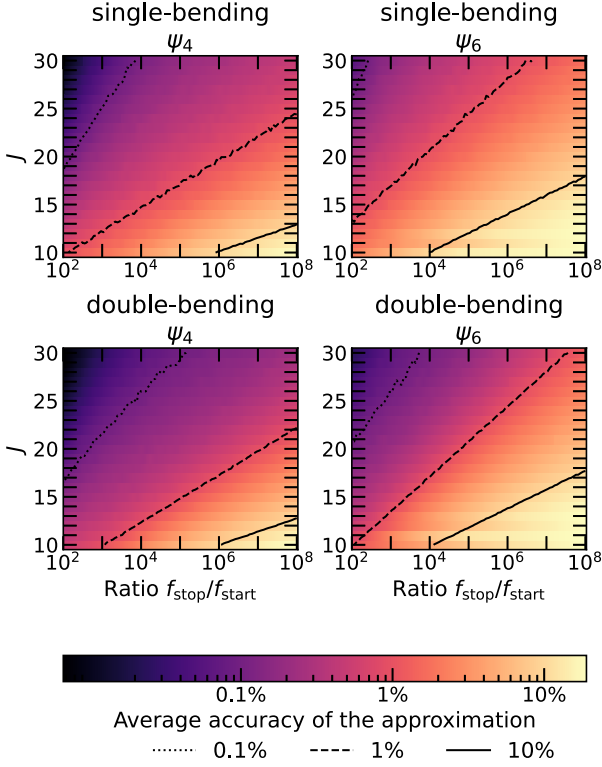
## APPENDIX B: QUALIFYING THE APPROXIMATION

### B1 Condition number

The condition number – using the  $L^2$  norm – associated with the linear system of equation (10) is computed as the ratio of the maximum and minimum eigenvalues. The condition number is plotted in Fig. B1 as a function of the size of the frequency grid for



**Figure B1.** Condition number of the linear system for the approximation as a function of the ratio between the maximal and minimal frequency. The basis functions  $\psi_4$  and  $\psi_6$  are plotted respectively with triangular markers on solid lines and circle markers on dashed-dotted lines. Colours present the number of basis functions for the approximation.



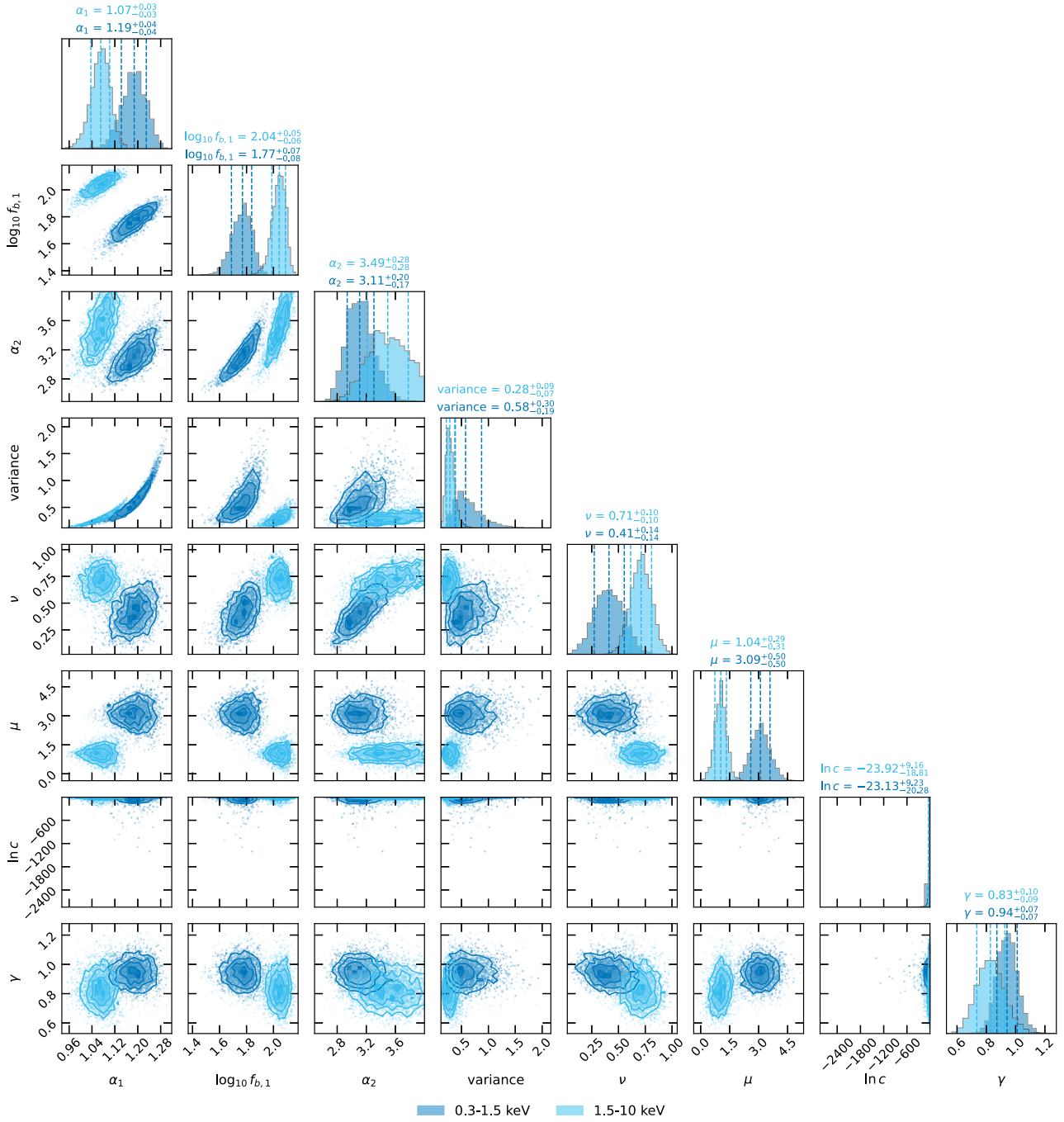
**Figure B2.** Average accuracy of the power spectrum approximation for  $\psi_4$  and  $\psi_6$  as a function  $J$  (number of basis functions) and size of the frequency grid. The top panels show the accuracy for the single-bending power-law model and the bottom panel show the accuracy for the double-bending power-law model. The accuracy increases for darker colours, the dotted, dashed, and solid lines represents respectively the level of 0.1 per cent, 1 per cent, and 10 per cent accuracy.

various values of  $J$ . We observe that it increases with the number of basis functions used and decreases when the frequency grid spans several orders of magnitudes.

### B2 Quality of the approximation

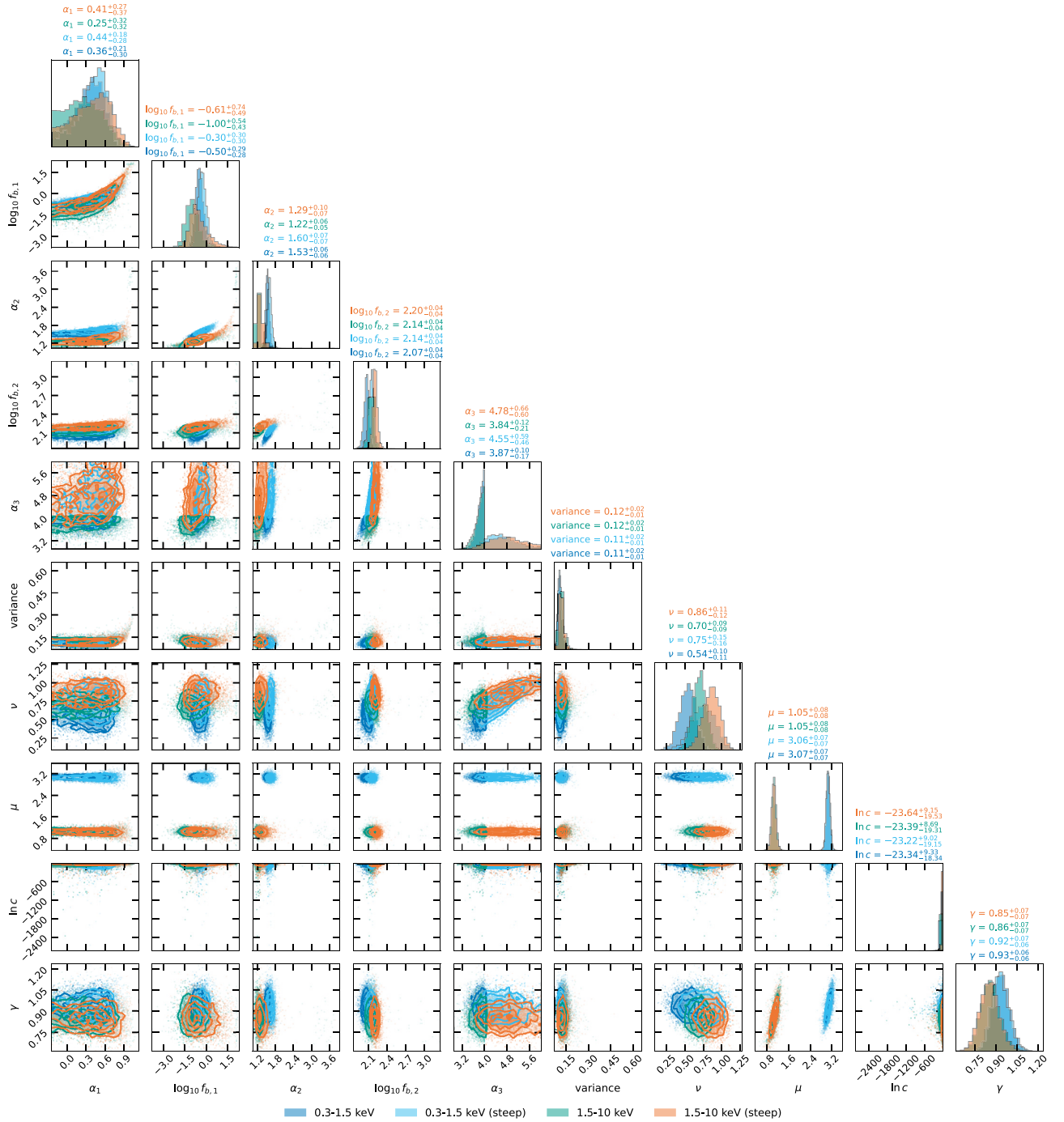
The quality of the power spectral approximation can be checked by computing the average normalized residual defined as  $\frac{1}{N} \sum_{i=1}^N \frac{|\mathcal{P}(f_i) - \tilde{\mathcal{P}}(f_i)|}{\mathcal{P}(f_i)}$  for multiple power spectral shapes approximated by  $J$  basis functions over a grid of frequencies given by  $f_{\text{stop}}/f_{\text{start}} = T/\Delta t/2$ . In Fig. B2, we show the accuracy of the approximation as a function of  $J$ ,  $f_{\text{stop}}/f_{\text{start}}$  and the model for  $\psi_4$  and  $\psi_6$ . We see that overall when choosing  $J$  between 20 and 30 the approximation has a good accuracy of about 1 per cent.

## APPENDIX C: POSTERIOR DISTRIBUTIONS



**Figure C1.** Posterior samples for the single bending power-law model in the soft and hard energy bands for the *XMM-Newton* and *Swift* observations of Ark 564.





**Figure C2.** Posterior samples for the double bending power-law model in the soft and hard energy bands for the *XMM-Newton* and *Swift* observations of Ark 564.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.